

Univerzita Karlova
1. lékařská fakulta

Studijní program: Doktorské studijní programy v biomedicině
Studijní obor: Molekulární a buněčná biologie, genetika a virologie



UNIVERZITA KARLOVA
1. lékařská fakulta

Mgr. Jan Hojný

**Analýza genových produktů vznikajících v důsledku alternativního sestřihu
pre-mRNA a jejich význam v onkogenezi karcinomu prsu.**

**Analysis of pre-mRNA alternative splicing products and their importance in
breast cancer oncogenesis.**

Dizertační práce

Školitel:
MUDr. Petra Kleiblová, Ph.D.

Praha, 2019

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 31. 5. 2019

Jan Hojný

Podpis:

Identifikační záznam:

HOJNÝ, Jan. Analýza genových produktů vznikajících v důsledku alternativního sestřihu pre-mRNA a jejich význam v onkogenezi karcinomu prsu. [*Analysis of pre-mRNA alternative splicing products and their importance in breast cancer oncogenesis*]. Praha, 2019. 108 s., 5 příl. Dizertační práce (Ph.D.). Univerzita Karlova, 1. lékařská fakulta, Ústav biochemie a experimentální onkologie. Vedoucí práce Kleiblová, Petra.

Poděkování:

Na tomto místě bych rád poděkoval především své školitelce MUDr. Petře Kleiblové Ph.D., dále prof. MUDr. Zdeňku Kleiblovi Ph.D. a ostatním členům našeho výzkumného týmu za pomoc, cenné rady, ochotu a hlavně trpělivost, kterou mi poskytli při vypracovávání této práce. Dále bych rád poděkoval Ing. Vajnerovi ze společnosti Sipoch spol. s r.o. za ochotu při zapůjčení a nastavení pipetovací stanice Gilson PipetMax, bez které by bylo pipetování téměř 20000 qPCR reakcí velmi obtížné. V neposlední řadě bych rád poděkoval své ženě, rodině a přátelům, za trpělivost a podporu a na závěr studentskému klubu Přírodovědecké fakulty UK Mrtvá Ryba a všem jeho návštěvníkům za významnou pomoc při obnově psychických sil v průběhu studia.

Práce na projektu byla podpořena granty IGA NT12280, SVV 260367

ABSTRAKT

Karcinom prsu je celosvětově nejčastěji diagnostikované nádorové onemocnění u žen. V 5-10 % všech případů je pozorována genetická souvislost, obvykle způsobená patogenní mutací v některém z predispozičních genů. Ačkoliv byla řada poškozujících mutací v kódující sekvenci těchto genů popsána, u velkého procenta familiárních případů (> 50 %) nebyla příčina dosud nalezena. Řada identifikovaných patogenních mutací byla lokalizována v konsenzních sestřihových místech, které mají za následek vznik aberantních sestřihových forem mRNA a z nich se odvíjející poškozené proteiny. Málo je však známo o variantách poškozujících regulační sestřihová místa, která mohou vést k tvorbě obdobných forem mRNA.

Pro nepřímou analýzu variant, ovlivňujících přirozený sestřih, jsme navrhli metodiku detekce sestřihových variant jakéhokoli genu založenou na multiplexní PCR a následné analýze pomocí NGS s vysokou citlivostí. Ověření této metodiky na modelu *BRCA1* odhalila přítomnost 94 sestřihových variant v leukocytech periferní krve, zdravé prsní a přilehlé tukové tkáni, čímž byl vytvořen dosud nejpodrobnější katalog fyziologicky se vyskytujících mRNA variant *BRCA1*.

Nejčastěji se vyskytující varianty, zachovávající čtecí rámec, byly přesně kvantifikovány pomocí RT-qPCR, která odhalila přítomnost 6 ubikvitně se vyskytujících alternativních transkriptů s relativní expresí > 1 % celkové exprese *BRCA1* ($\Delta 5$; $\Delta 9_{10}$; $\Delta 9_{10,11q}$; $\blacktriangledown 13$ a IRIS). Dále jsme prokázali tkáňově specifickou míru exprese u variant $\Delta 9_{10}$, $\blacktriangledown 13$ a IRIS. Většina ubikvitních variant si pravděpodobně zachovává charakter wild-type formy, či vykazuje dosud neobjasněnou regulační funkci.

Výsledky práce objasňují složení a množství mRNA variant *BRCA1* v relevantních zdravých tkáních. Na základě tohoto katalogu je např. možné okamžitě identifikovat aberantní sestřihové mRNA varianty, vyskytující se v nádorové tkáni, či prokázat přítomnost mutace, vedoucí k deregulaci sestřihu pre-mRNA, v případě negativního výsledku mutační analýzy *BRCA1*.

Klíčová slova: Alternativní sestřih, *BRCA1*, karcinom prsu, analýza genové exprese

ABSTRACT

Breast cancer is the most common tumor disease diagnosed in women worldwide. The hereditary character of this disease is observed in 5-10 % of all cases, and it is usually caused by a pathogenic mutation in one of the predisposition genes. Although a variety of pathogenic mutations in the coding sequences of these genes was described, the cause of the disease is still unknown in many familial cases (> 50%). A great number of identified pathogenic mutations were localized in the consensus splicing sites, which results in the formation of aberrant mRNA splicing variants and their damaged protein isoforms. However, little is known about mutations affecting regulatory splicing sites, which can result in the translation of similarly affected mRNAs.

In this work, we proposed a method for indirect detection of mutations affecting the natural splicing pattern of any gene of our interest based on multiplex PCR and NGS with high sensitivity. Verification of this method on the BRCA1 model gene revealed the presence of the total of 94 splicing variants in peripheral leucocytes and healthy breast and adjacent fat tissues. This is the most detailed catalogue of physically occurring BRCA1 mRNA variants thus far.

The most commonly occurring variants, maintaining open reading frame, were quantified by RT-qPCR which resulted in the characterization of 6 ubiquitously expressed alternative splicing variants with a relative expression > 1 % of total BRCA1 ($\Delta 5$; $\Delta 9_{-10}$; $\Delta 9_{-10,11q}$; $\nabla 13$ and IRIS). Furthermore, we detected tissue specific levels in the expression of $\Delta 9_{-10}$, $\nabla 13$ and IRIS variants. The majority of ubiquitous variants probably result in protein isoforms which maintain the BRCA1 wild-type character or an unknown (probably regulatory) function.

This study fully reveals the qualitative and quantitative splicing pattern of BRCA1 mRNA variants in relevant healthy human tissues. Based on this, we can instantly detect aberrantly spliced BRCA1 mRNA variants e.g. in tumor tissue and reveal the presence of mutation affecting the regulatory splicing site in cases of negative mutation analysis of *BRCA1*.

Key words: Alternative splicing, BRCA1, breast cancer, gene expression analysis

OBSAH

Abstrakt	5
Abstract.....	6
Obsah.....	7
1 Úvod.....	9
1.1 Sestřih prekurzorové mRNA	11
1.1.1 Mechanismus sestřihu	12
1.2 Alternativní RNA sestřih.....	14
1.2.1 Význam alternativního sestřihu.....	15
1.2.2 Mechanismus alternativního sestřihu.....	15
1.2.2.1 Regulační sestřihové proteiny.....	16
1.2.2.2 Regulace spojená s transkripcí.....	16
1.2.2.3 Regulace spojená s rychlostí RNA polymerázy II.....	17
1.2.2.4 Regulace spojená se strukturou chromatinu	18
1.2.2.5 Alternativní sestřih na základě volných sestřihových míst.....	20
1.3 Význam sestřihu u genů predisponujících k dědičné formě karcinomu prsu.....	21
1.3.1 Predispoziční geny dědičné formy karcinomu prsu	21
1.3.2 BRCA1	22
1.3.2.1 Struktura transkriptu BRCA1	23
1.3.2.2 Funkce proteinu BRCA1	25
1.3.3 Alternativní sestřihové varianty BRCA1 a jejich význam	26
2 Cíle práce.....	31
3 Materiál a metody.....	32
3.1 Strategie řešení	32
3.2 Charakteristika vyšetřovaného souboru	33
3.2.1 Biologické vzorky od pacientek s karcinomem prsu a kontrol	33
3.2.2 Buněčné linie.....	33
3.3 Odběr, izolace a zpracování biologického materiálu	34
3.3.1 Odběr vzorků.....	34
3.3.2 Izolace celkové RNA z buněčných linií	34
3.3.3 Izolace celkové RNA z periferní krve.....	34
3.3.4 Izolace celkové RNA ze solidních tkání	35
3.3.5 Kontrola kvality izolované RNA.....	36
3.3.6 Syntéza cDNA.....	37
3.3.7 PCR kontrola reverzní transkripce syntetizované cDNA pro identifikaci ASV	38
3.4 NGS identifikace alternativních sestřihových variant BRCA1	39
3.4.1 Multiplexní PCR	40
3.4.1.1 Návrh multiplexní PCR (mPCR)	40

3.4.1.2	Vlastní mPCR reakce.....	42
3.4.2	Velikostní selekce	44
3.4.3	Příprava sekvenační knihovny	45
3.4.4	Sekvenování na platformě MiSeq	46
3.4.5	Biostatistické zpracování dat.....	47
3.5	Kvantifikace exprese vybraných nalezených ASV BRCA1 pomocí RT-qPCR	49
3.5.1	Návrh primerů pro RT-qPCR analýzu	49
3.5.2	Preamplifikace.....	50
3.5.3	RT-qPCR analýza a analýza křivek tání.....	51
3.5.4	Stanovení efektivity qPCR reakcí	52
3.5.5	Kvantifikace změn exprese jednotlivých ASV a jejich statistické vyhodnocení.....	52
4	Výsledky.....	54
4.1	Izolace a zpracování biologického materiálu	54
4.1.1	Izolace a kontrola kvality celkové RNA z lidských tkání	54
4.1.2	Optimalizace reverzní transkripce.....	54
4.2	Identifikace alternativních sestřihových variant BRCA1 v biologických vzorcích	56
4.2.1	Kontrola správného průběhu syntézy cDNA pro identifikaci ASV	56
4.2.2	Optimalizace primerů pro multiplexní PCR reakce	57
4.2.3	Optimalizace průběhu multiplexních PCR reakcí	58
4.2.4	Optimalizace velikostní selekce produktů mPCR reakcí	59
4.2.5	Souhrnné výsledky mapování ASV BRCA1	60
4.2.6	Tkáňová specifická nalezených variant.....	64
4.2.7	Kvantitativní posouzení variant s volnými sestřihovými místy	65
4.3	Kvantifikace vybraných ASV v biologických vzorcích pomocí RT-qPCR.....	69
4.3.1	Analýzovaný soubor vzorků.....	69
4.3.2	Optimalizace qPCR reakce.....	69
4.3.3	Zhodnocení vybraných referenčních genů	70
4.3.4	Celková exprese BRCA1 ve vyšetřovaných tkáních.....	72
4.3.5	Relativní exprese ASV BRCA1	74
4.3.5.1	Tkáňově specifická exprese ASV BRCA1	76
4.3.5.2	Relativní exprese ASV BRCA1 ve skupině zdravých kontrol a pacientek s karcinomem prsu	80
4.3.6	Laterálně specifická exprese jednotlivých variant	83
5	Diskuze.....	87
6	Závěr.....	97
7	Seznam použitých zkratk.....	99
8	Seznam použité literatury	100
	Přílohy	109

1 ÚVOD

Podle současného, obecně přijímaného modelu kancerogeneze, je pro rozvoj maligní transformace nutné nakumulovat několik konsekventních mutací v regionech genů regulujících základní buněčné procesy (Cahill et al., 1999). Úspěšná nádorová buňka musí pozměnit svoji buněčnou fyziologii, aby byla schopna autonomní proliferace, potlačení apoptózy, rezistence k signálům růstových inhibitorů, neomezené možnosti replikačního potenciálu a následně i možné angiogeneze, invaze a metastazování (Hanahan a Weinberg, 2000).

V průběhu kancerogeneze jsou tradičními cíli mutací tzv. tumor supresorové geny, jejichž proteinové produkty jsou zapojeny v protinádorových procesech, jako jsou opravy DNA, zastavení buněčného cyklu, iniciace apoptózy atd. Avšak dle Knudsonovy teorie dvou zásahů je nutné alterovat sekvenci DNA na obou alelách genu, aby byla vyřazena funkce příslušného tumor supresorového proteinu (Knudson, 1971). Většina nádorových onemocnění proto vzniká až v pokročilejší fázi života, ve chvíli, kdy se v buňce nakumuluje dostatečné množství somaticky vzniklých mutací nejen v tumor supresorových genech. Existuje však řada nádorových onemocnění, nebo jejich forem, které se projevují ve výrazně nižším věku. U mnoha z nich jsou predispozice k těmto onemocněním v rodině děděny. Důvodem je výskyt zárodečné mutace v jednom z tumor supresorových genů, čímž je oslabena protinádorová „bariéra“ a usnadněn proces vzniku tumorogeneze. Příkladem mohou být příčinné mutace v *BRCA1*, které ve většině případů způsobují dědičnou formu karcinomu prsu.

Identifikace příčinných mutací u dědičných nádorových onemocnění jsou v dnešní době prováděny pomocí masivního paralelního sekvenování, tzv. sekvenování nové generace – NGS (New Generation Sequencing), které nám umožňuje analyzovat genetickou informaci u rozsáhlých genových panelů, případně celých exomů či genomů jedinců. Stinnou stránkou takto mocného nástroje je detekce velkého množství často neznámých alterací s nejasným dopadem (VUS – variant of uncertain significance), které je obtížné rychle charakterizovat. Kromě relativně dobře popsanych změn v kódujících a přilehlých oblastech (bodové mutace, mutace konsenzních sestríhových míst, delece/inzerce nebo velké chromozomální přestavby) nebo epigenetických změn ovlivňujících expresi daného genu (hypermetylace tzv. CpG ostrůvků v promotorové oblasti genu, modifikace histonů a změna oblasti na transkripčně neaktivní heterochromatin), existuje celá řada mutací postihujících intronové oblasti, o jejichž dopadu není dosud mnoho známo (Scholzova et al., 2007).

Pokud intronové mutace postihnou sekvenci, na kterou nasedají sestřihové nebo regulační faktory (ať už pozitivně nebo negativně působící), může dojít ke změně sestřihového vzorce pro dané geny, a tak k ovlivnění produktů genové exprese. Kvalitativní (i kvantitativní) změna na úrovni genové exprese může mít v řadě případů velice závažné důsledky. Například tvorba alternativních sestřihových variant (ASV) androgenního receptoru je úzce spjata se vznikem a vývojem metastatických karcinomů prostaty (Paschalis et al., 2018).

Ačkoliv se studium intronových sekvencí po nástupu NGS mnohonásobně zefektivnilo, charakterizovat dopad nalezených intronových odchylek od referenčního genomu u daného jednovlivce je vzhledem k obrovskému množství dat z celogenomových studií a k vysoké heterogenitě intronových sekvencí napříč populacemi a značnému množství privátních variant jednovlivce, stále velmi obtížné. Ke správné interpretaci mutací, postihujících regulační sestřihové sekvence, je nezbytné tyto sekvence napříč introny identifikovat. I přes existenci řady různých prediktivních algoritmů *in-silico* detekce regulačních míst, není identifikace těchto oblastí zcela spolehlivá.

Alternativní, v této práci navrhovaný, přístup pro odhalení variant postihujících regulační místa sestřihu, vychází ze znalosti všech fyziologicky vznikajících sestřihových forem mRNA. Detekce různých sestřihových variant je v současných podmínkách RNA sekvenování nové generace (RNA-Seq) nepoměrně snazší, a proto je i relativně snadné touto metodou odhalit aberantní sestřihovou variantu mRNA při znalosti souboru fyziologicky se vyskytujících sestřihových variant.

Předkládaná práce je proto zaměřena na problematiku vzniku fyziologicky se vyskytujících alternativních sestřihových variant, návrhu metodiky pro jejich komplexní identifikaci a kvantifikaci s vysokou přesností na modelu *BRCAl*, což je první a nezbytný krok pro odlišení fyziologicky se vyskytujících variant od variant vzniklých na základě poškození sestřihového vzorce na úrovni mutací DNA.

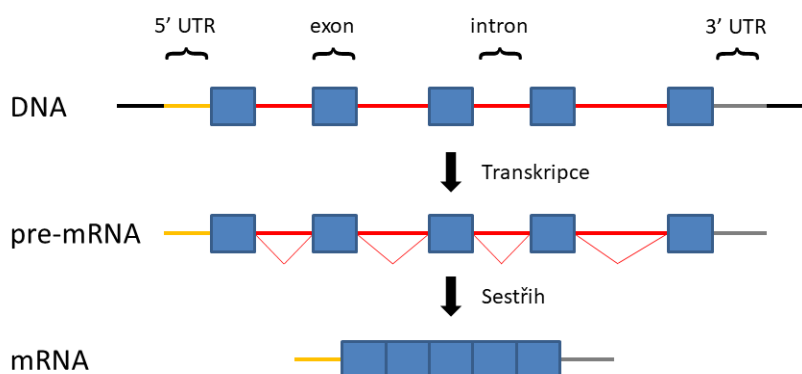
Přes veškerý pokrok a znalosti tyto informace (kromě sporadických analýz) u většiny genů chybí.

1.1 Sestřih prekurzorové mRNA

Eukaryotické geny jsou sestaveny ze dvou základních funkčních struktur, které se nazývají exony a introny. Jediný rozdíl mezi těmito úseky genomové DNA je v jejich přepisu do mediátorové RNA (mRNA; messenger RNA), a tedy přítomnost ve finální podobě mRNA. Přepsané exony v mRNA zůstanou a slouží jako informace pro překlad do proteinů, zatímco introny jsou vystřiženy a degradovány.

V první fázi genové exprese je přepsána DNA celého genu do RNA a tím vzniká tzv. primární transkript neboli prekurzorová mRNA (pre-mRNA). Pre-mRNA je dále upravována několika specifickými kroky, mezi které patří vytvoření 5' čepičky, sestřih a syntéza poly-A konce. Zatímco 5' čepička a poly-A konec mají zásadní roli pro stabilitu maturované mRNA, na sestřihu závisí složení kódující sekvence. Teprve po dokončení těchto základních úprav vzniká finální produkt – mRNA (Alberts et al., 2008).

U genů s více než jedním exonem, kterých je v lidském genomu většina, je potřeba mít pro překlad z mRNA do proteinové sekvence jednotnou kódující sekvenci zbavenou vmezeřených intronů, které často zaujímají až 95 % sekvence genu (Mattick a Gagen 2001). Proces odstranění intronových sekvencí se nazývá sestřih pre-mRNA. Sestřih pre-mRNA začíná ve většině případů ihned po nasyntetizování prvního úseku (intronu) určeného k eliminaci (sestřižení), probíhá kotranskripčně. Produktem je celistvá kódující sekvence definující čtecí rámec pro syntézu příslušného proteinu, která je obklopená přilehlými 5' a 3' nepřekládanými oblastmi (UTR; untranslated region; Obr. 1; Pollard a Earnshaw, 2008; Alberts et al., 2008).



Obrázek 1: Schéma vzniku maturované mRNA přepisem z genu a sestřihem. Modré čtverce reprezentují kódující sekvenci exonů, červené úseky introny, žlutý (5') a šedý (3') úsek znázorňuje nepřekládané oblasti (UTR) exonů a černě je znázorněna přilehlá DNA.

1.1.1 Mechanismus sestřihu

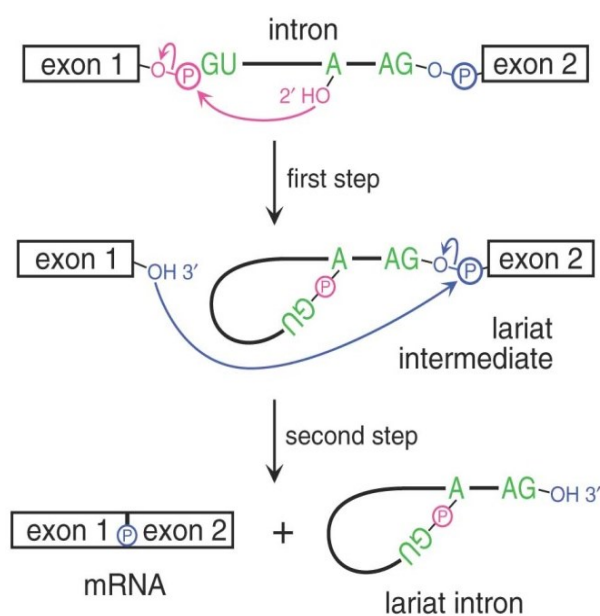
Sestřih pre-mRNA je v eukaryontních buňkách prováděn výlučně za pomoci tzv. spliceosomu, což je multiproteinový komplex specificky zajišťující tento kriticky důležitý proces maturace mRNA. Spliceosom obsahuje přes 200 proteinů, které vzájemnými interakcemi zajišťují dostatečnou komplexitu a flexibilitu sestřihového systému. Pomocí spliceosomu je dle obecného klíče rychle a velmi přesně rozeznáváno a vystřihováno velké množství různých intronů, vyskytujících se napříč eukaryotickým genomem (Alberts et al., 2008).

Katalytické jádro spliceosomu tvoří pět hlavních sestřihových faktorů – U1, U2, U4, U5 a U6. Tyto faktory jsou označovány jako malé ribonukleoproteiny (snRNP; small ribonucleoprotein particles), protože obsahují krátké jaderné RNA sekvence (snRNA; small nuclear RNA). RNA součásti snRNP slouží k rozpoznávání DNA sekvencí důležitých pro sestřihové pochody. Ribonukleoproteiny se na nově syntetizovanou pre-mRNA váží ko-transkripčně (v průběhu transkripce) a společně zajišťují základní funkce sestřihu, které jsou: a) rozeznávání hlavních (na začátku/konci intronů) a pomocných (místo větvení v intronu) sestřihových míst, b) zajišťování vhodných podmínek pro transesterifikační reakce vlastní sestřihové události pomocí vzájemné kolokalizace (Alberts et al., 2008; Pollard a Earnshaw, 2008).

Přiblížení a spojení dvou exonů za současného vystřihnutí intronového úseku, ležícího mezi těmito exony ve formě smyčky (lariátu), vyžaduje významné dynamické změny ve strukturách sestřihového aparátu (Alberts et al., 2008).

- V první fázi sestřihové události rozeznává ribonukleoprotein U1 5' sestřihové místo, protein BBP (branch-point binding protein; vazebný protein na místo větvení) rozeznává místo větvení a U2AF (U2 auxiliary factor; pomocný faktor U2) místo polypyrimidinového traktu.
- V druhé fázi nahrazuje BBP v místě větvení ribonukleoprotein U2 a vytvoří komplex s U2AF.
- V průběhu třetí fáze dojde k přestavbě celého komplexu po vstupu tří snRNP (U4/U6 a U5), jenž vyústí v přiblížení U1 a U2 navázané na pre-mRNA. Přiblížení U1 a U2 umožní první transesterifikační reakci mezi 5' koncem intronu a adenosinem v místě větvení za vytvoření smyčky.

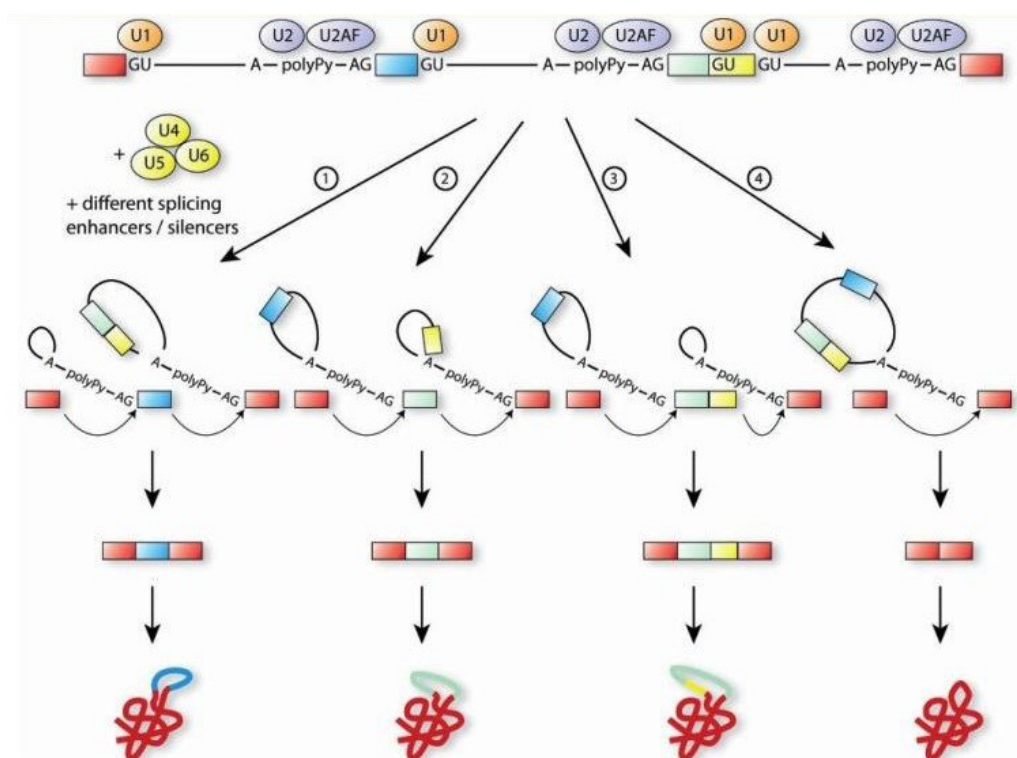
- Ve čtvrté fázi dochází opět k přestavbě komplexu. Rozpadne se U4/U6 heterodimer, což umožní výměnu U1 v 5' sestřihovém místě za U6. Tak vzniká aktivní místo pro druhou transesterifikační reakci. Touto reakcí, která spojí konce exonů, je sestřihová událost ukončena (Obr. 2; Valadkhan et al., 2003; Alberts et al., 2008).



Obrázek 2: Schéma obou transesterifikačních reakcí během sestřihové události. Zeleně jsou označeny 5' (GU) a 3' (AG) konsensní sestřihová místa v intronu a adenosin v místě větvení. **Růžově** jsou zobrazeny skupiny, které se účastní první reakce (první krok) a **modře** skupiny, účastníci se druhé reakce (druhý krok). Hydroxylová skupina (2'-OH) adenosinu v místě větvení v prvním kroku atakuje fosfodiesterovou vazbu v 5' exon-intronového spojení (5' sestřihové místo). Tím se přeruší fosfodiesterová vazba mezi 5' exonem a intronem a dojde k vytvoření nové fosfodiesterové vazby mezi 5' koncem intronu a adenosinem místa větvení (5'-2') za vzniku smyčky (lariátu). Ve druhém kroku je fosfát ve spoji 3' konce intronu a 5' konce exonu (3' sestřihové místo) atakován volnou 3'-OH skupinou exonu předchozího a dojde ke spojení exonů za uvolnění smyčky. Smyčka je později degradována. Obrázek převzat z Horowitz, 2012.

1.2 Alternativní RNA sestřih

Kromě konstitutivního (kanonického) sestřihu, kterým označujeme sestřih vedoucí ke vzniku referenčního / kanonického / resp. divokého typu mRNA (wt; wild-type), dochází za určitých okolností i k sestřihu alternativnímu, a to až u 95 % multiexonových genů lidského genomu (Pan et al., 2008). Jeho podstatou je rozdílné rozpoznání cis-regulačních sekvencí pre-mRNA sestřihovým aparátem na základě vlivu řady regulačních faktorů. To má za následek vznik odlišné (alternativní) mRNA struktury při zachování intaktní primární struktury genové DNA¹ (Obr. 3; Alberts et al., 2008; Pollard a Earnshaw, 2008).



Obrázek 3: Schéma dopadu alternativního sestřihu na proteinový produkt. Červenými obdélníky jsou znázorněny konstitutivní exonové sekvence. Modrými, zelenými a žlutými obdélníky jsou vyznačeny alternativní exony. Intronové sekvence s 5' sestřihovým místem (GU), místem větvení (A), polypyrimidinovým místem (polyPY) a 3' místem větvení (AG) jsou znázorněny černě. Sestřihové faktory U1 – U6 a U2AF jsou vyznačeny oranžovými, fialovými a žlutými elipsami. Šipkami s čísly 1-4 jsou znázorněny různé situace alternativního sestřihu, které vytvářejí různé produkty na základě tkáňové specifičnosti nebo různého působení pomocných regulačních sestřihových faktorů. Obrázek převzat z Wood et al., 2010.

¹ Pokud by se však na úrovni DNA vyskytla mutace postihující cis-regulační sestřihovou sekvenci DNA, a na základě této mutace by došlo k patologickým změnám v sestřihu, vzniklá sestřihová varianta by byla označena jako patogenní, a jednalo by se o sestřih aberantní.

1.2.1 Význam alternativního sestřihu

S nástupem NGS a pokročilé bioinformatiky se snížilo množství odhadovaných protein kódujících genů v lidském genomu z 25 až 40 tisíc (90. léta) na recentně udávaných 19 tisíc (Ezkurdia et al., 2014). Vysoká variabilita proteomu při překvapivě nízkém množství zdrojových genů je z velké míry způsobena alternativním sestřihem, kdy z jednoho genu vzniká obvykle několik proteinových izoforem s různorodými vlastnostmi. Ačkoliv je fenomén alternativního sestřihu znám už více než 30 let, je mu v poslední době, nejen z tohoto důvodu, věnována zvýšená pozornost (Pollard a Earnshaw, 2008; Kornblihtt et al., 2013).

Finální proteinový produkt může být alternativním sestřihem ovlivňován kvalitativně i kvantitativně. Mezi kvalitativní ovlivnění patří výpadek exonu(ů) či inserce alternativního exonu(ů) či vzácné vytvoření předčasného stop kodonu v alternativním exon-exonovém spojení. Mezi kvantitativní ovlivnění patří např. změny v 5'- či 3'-UTR, které ovlivňují transport mRNA, stabilitu mRNA či sílu translace a další.

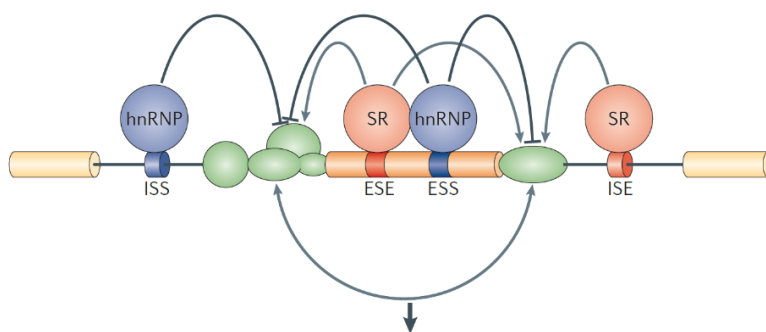
Translací alternativně sestřižených mRNA mohou vznikat soubory proteinů obsahující různé kombinace funkčních domén (kódovaných různými exony) z jediné pre-mRNA (Obr. 3). Alternativní produkt se tak může od kanonického (wt) produktu lišit strukturně, funkčně (enzymatickou nebo signální aktivitou), asociacemi s vazebnými partnery, intracelulární lokalizací nebo stabilitou (Alberts et al., 2008; Pollard a Earnshaw, 2008).

1.2.2 Mechanismus alternativního sestřihu

Hlavním principem alternativního sestřihu je kompetice „silných“ a „slabých“ konsensních sestřihových míst, které jsou přítomny napříč pre-mRNA, o komponenty spliceosomu (viz kapitola 1.1.1. Mechanismus sestřihu). „Silnými“ sestřihovými místy označujeme ta, která vytváří díky své sekvenci lepší vazebné podmínky, a tedy i afinitu pro sestřihové faktory, a naopak „slabými“ ta, kde je vazebná afinita nižší. O tom, zda bude sestřihové místo právě pro konkrétní sestřihovou událost využito, dále rozhodují pozitivní (enhancery) a negativní (silencery) regulační sekvence sestřihu společně s navázanými regulačními proteiny, aktuální strukturou chromatinu a rychlostí transkripce DNA-dependentní RNA polymerázy II (RNAP II) a dalšími (Alberts et al., 2008; Pollard a Earnshaw, 2008; Kornblihtt et al., 2013).

1.2.2.1 Regulační sestřihové proteiny

Regulačních proteinů sestřihu známe několik tříd. První ze dvou hlavních tříd jsou faktory bohaté na serin a arginin (SR – serine-arginine rich) a jejich příbuzné proteiny (SR-related proteins; Krecic a Swanson, 1999; Cowper et al., 2001). Druhou třídou jsou heterogenní nukleární ribonukleoproteiny (hnRNP; Fu, 1995). Dalšími faktory, ovlivňujícími sestřih, mohou být tkáňově specifické faktory jako např. nPTB a PTB, NOVA a FOX (Kafasla et al., 2012; Jelen et al., 2007; Lee et al., 2009). Všechny tyto regulační elementy se váží na pre-mRNA do konkrétních sestřihových míst lokalizovaných uvnitř exonů (ESE – exon splicing enhancer / ESS – exon splicing silencer) či intronů (ISE – intron splicing enhancer / ISS – intron splicing silencer). SR proteiny a jejich příbuzná skupina se váže na *cis* elementy, které zesilují sestřih mezi nejbližším 5'a 3' sestřihovým místem, zatímco hnRNP faktory se váží na tlumící sekvence u 3' místa, kde inhibují možnost sestřihu, nebo podporují použití vzdálenějšího 5' sestřihového místa (Shen a Green, 2004; Blencowe, 2000; Fairbrother a Chasin, 2000). Poměr jaderné koncentrace pozitivně působících SR proteinů a inhibičních hnRNP faktorů je proto jedním ze základních regulačních mechanismů alternativního sestřihu (Caceres et al., 1994; Obr 4.)



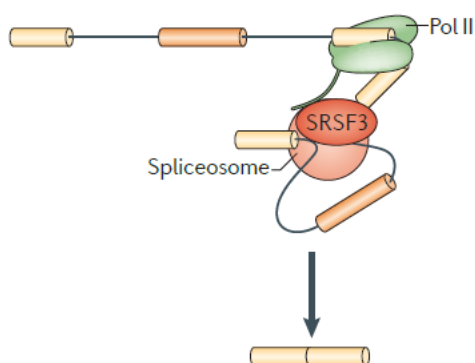
Obrázek 4: Schéma regulačních sekvencí a faktorů alternativního sestřihu. Modré kruhy znázorňují proteiny rodiny heterogenních nukleárních ribonukleoproteinů (hnRNP). **Červené kruhy** znázorňují faktory rodiny serin-arginin bohatých proteinů (SR). **Modré válce** znázorňují intron (ISS) a exon (ESS) negativně regulační oblasti. **Červené válce** znázorňují exonové (ESE) a intronové (ISE) regulační oblasti. **Zeleně** jsou znázorněny proteiny přímo se účastnící sestřihu, **žlutě** exony a **modrou čarou** intronové oblasti pre-mRNA. Obrázek převzat z Kornblihtt et al., 2013.

1.2.2.2 Regulace spojená s transkripcí

Původní představa o tvorbě a maturaci mRNA byla rozdělena do jednotlivých, vzájemně nezávislých kroků – nejprve translace pre-mRNA a až po jejím ukončení sestřih pre-mRNA –

a to i navzdory teoriím, které popisovaly sestřih kotranskripčně (Beyer a Osheim, 1988). Teprve v posledním desetiletí se definitivně prokázalo, že kotranskripčně probíhá většina sestřihových událostí, a to napříč eukaryotickými organizmy (včetně člověka; Ramanouskaya a Grinev, 2017).

Důležitým objevem byl mechanismus umístění sestřihových faktorů do míst transkripce pomocí vazby na proteinový transkripční superkomplex (obsahující RNAP II, základní transkripční faktory, tkáňově / genově specifické transkripční faktory, pomocné proteiny, zesilovače transkripce a další), ve kterém má zásadní úlohu C terminální doména (CTD) RNAP II (Muñoz et al., 2010). CTD RNAP II je schopna vázat řadu alternativních sestřihových faktorů (SR proteinů), které kotranskripčně svojí přítomností ovlivňují mechanismus alternativního sestřihu (Das et al., 2007). Přímý vztah mezi vazbou SR proteinů na CTD RNAP II a alternativním sestřihem byl prokázán, mimo jiné, in-vitro pokusy se zkrácenou CTD RNAP II a inaktivací genu pro SR protein SRSF3 (Obr. 5; de la Mata a Kornblihtt, 2006).

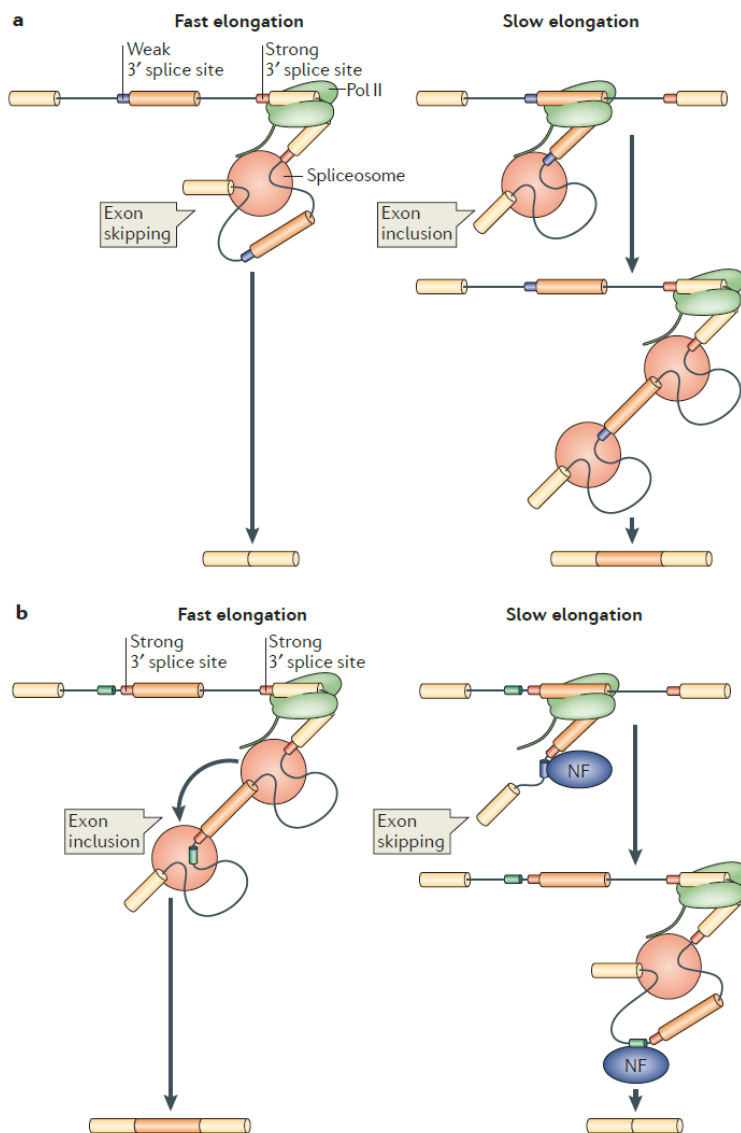


Obrázek 5: Schéma regulace alternativního sestřihu vazbou regulačních proteinů na RNAP II. Tmavě oranžově je označen alternativní exon 33, který je na základě přítomnosti negativně ovlivňujícího SR proteinu SRSF3 (červeně) navázaném na CTD RNAP II (zeleně) vystřižen z pre-mRNA fibronektinu. Obrázek převzat z Kornblihtt et al., 2013.

1.2.2.3 Regulace spojená s rychlostí RNA polymerázy II

Dalším, dlouhou dobu diskutovaným, mechanismem, ovlivňujícím alternativní sestřih, je vliv rychlosti elongace RNAP II (Aebi a Weissmann; 1987). Původní teorie založená na *in vitro* sestřihových pokusech vysvětluje ovlivnění alternativního sestřihu na základě vzniku sekundárních struktur RNA, jejichž dynamika vzniku je ovlivněna rychlostí RNAP II (Eperon et al., 1988). Tato teorie byla během let rozvíjena a postupně byl prokázán vliv promotoru na fosforylaci CTD RNAP II, a tím na rychlost její elongace (Cramer et al., 1997 a 1999; Kornblihtt, 2005). Principem tohoto mechanismu při „rychlé“ elongaci RNAP II je „přeskočení“ (vystřižení) alternativního exonu se slabým sestřihovým místem nebo

neumožnění nasednutí negativně působících faktorů na sestřih. Při elongaci „pomale“ je možné zahrnutí alternativního exonu nebo vystřížení konstitutivního exonu v důsledku negativního ovlivnění regulačních proteinů (Obr. 6; Luco et al., 2011 a Kornblihtt et al., 2013).



Obrázek 6: Schéma vlivu rychlosti elongace RNAP II na alternativní sestřih:

a) alternativní exon (tmavě oranžový válec) se slabým 3' sestřihovým místem je při rychlé elongaci (**Fast elongation**) vystřížen z pre-mRNA (**Exon skipping**) a při pomalé elongaci (**Slow elongation**) je v pre-mRNA ponechán (**Exon inclusion**)

b) konstitutivní exon se silným 3' sestřihovým místem je při rychlé elongaci zařazen, zatímco při pomalé elongaci se na regulační místo sestřihu (**zelený válec**) váže negativně působící jaderný faktor (**modře**) a exon je vystřížen. Obrázek převzat z Kornblihtt et al., 2013.

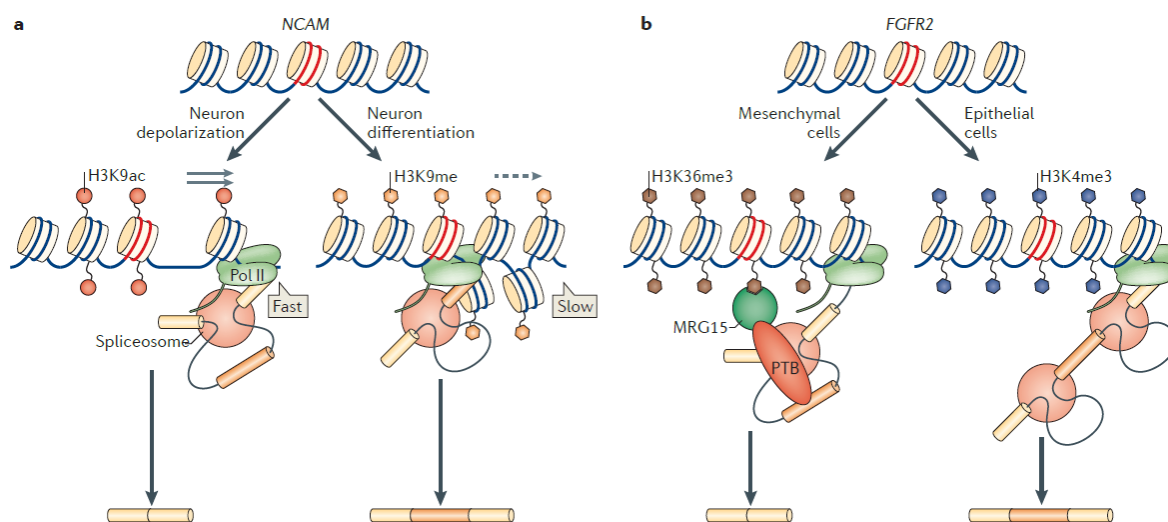
1.2.2.4 Regulace spojená se strukturou chromatinu

Po objevení základních regulačních mechanismů a během jejich intenzivního zkoumání začalo být jasné, že regulace alternativního sestřihu bude mnohem komplexnější, než se původně předpokládalo. Ani vazba RNA regulačních sestřihových proteinů či rychlost elongace RNAP II plně nevysvětlovaly všechny dosud pozorované aspekty alternativního sestřihu

in vivo. Teprve objevení vlivu struktury chromatinu a epigenetických modifikací histonů celkový model regulace alternativního sestřihu doplnilo (Luco et al., 2011).

Několik studií postupně prokázalo negativní vliv kompaktnosti chromatinu na rychlost RNAP II poté, co byla pozorována souvislost mezi mRNA transkripty se zahrnutým alternativním exonem a nízkou mírou replikace templátové DNA způsobenou vysokou mírou uspořádanosti chromatinu (Kadener et al., 2001). Byla ukázána schopnost remodelačního komplexu SWI/SNF ve spojení s CD44 vázat sestřihový faktor SAM68, a tím regulovat alternativní sestřih vytvořením fyzického bloku elongace RNAP II, která je takto zpomalena (Batsché et al., 2005).

Dále byl prokázán vliv post-translačních histonových modifikací na rychlost elongace RNAP II, který zároveň odpovídá míře transkripce daného genu. Zjednodušeně řečeno, modifikace podporující transkripci (např. H3K36me3, H3K4me2, H3K4me3 a H3K9ac) rozvolňují chromatin, usnadňují elongaci RNAP II, čímž ovlivňují sestřih (např. nezařazením alternativního exonu do finálního transkriptu). Naopak modifikace negativně podporující transkripci (např. H3K9me2, H3K9me3 a H3K27me3) působí na alternativní sestřih ve smyslu opačném (Kornblihtt et al., 2013, Obr. 7).



Obrázek 7: Schéma vlivu chromatinu na mechanismus alternativního sestřihu: a) příklad vlivu post-translační pozitivní (H3K9ac, vlevo) a negativní (H3K9me, vpravo) modifikace histonů (světle žlutě) na rychlost RNAP II (zeleně) a zahrnutí alternativního exonu (tmavě oranžově) do transkriptu genu *NCAM* b) vliv vazby komplexu MRG15 (zeleně) a sestřihového faktoru PTB (červeně) s modifikovaným H3K36me3 na vystřihnutí alternativního exonu (tmavě oranžově) genu *FGFR2* oproti modifikovanému H3K4me3, který komplex s MRG15 a PTB neposkytuje. Obrázek převzat z Kornblihtt et al., 2013.

Kromě obecného efektu post-translačních modifikací histonových konců byl pozorován vliv konkrétních modifikací na sestřihové události. Například trimetylované lysiny 36 na histonech H3 (H3K36me3) jsou vazebným cílem adaptorového proteinu MRG15, který interaguje se sestřihovým faktorem PTB. Tento komplex pak redukuje zahrnutí alternativního exonu do transkriptu. Modifikace H3K36me3 podporuje vznik transkriptu obsahující pouze kanonické exony, zatímco jiné pozitivně-transkripční modifikace tyto vazby neumožňují, což může napomáhat vzniku transkriptů alternativních (Obr. 7b; Kolasinska-Zwierz et al., 2009).

1.2.2.5 Alternativní sestřih na základě volných sestřihových míst

Zvláštním případem alternativního sestřihu jsou tzv. volná sestřihová místa (Leaky splice sites – LSS, někdy též nazýváno jako noisy splicing). Jedná se o konsenzní sestřihové sekvence, které si proteiny sestřihového aparátu vyberou v dané lokaci jako alternativu ke klasickému sestřihovému místu, a to nejen v případě přítomné mutace v klasickém sestřihovém místě. Dojde tak k tvorbě alternativního donorového nebo akceptorového sestřihového místa (SDS / SAS – splice donor / acceptor site).

S preferencí volného sestřihového místa proteiny spliceosomu je, kromě faktorů diskutovaných v předchozích kapitolách, spojena hlavně „síla“ (motiv) dané konsenzní sestřihové sekvence (v porovnání se „sílou“ sousedícího klasického sestřihového místa) a také chybovost sestřihového aparátu, která je odhadována na 0,7 % (Pickrell et al., 2010; Cooper and Mattox 1997).

Zajímavostí je čtenější využití volných sestřihových míst u méně exprimovaných genů, zatímco u vysoce exprimovaných genů, které častěji obsahují krátké introny, se s tímto fenoménem téměř nesetkáme (Pickrell et al., 2010).

1.3 Význam sestřihu u genů predisponujících k dědičné formě karcinomu prsu

Regulační mechanismy ovlivňující vznik alternativních sestřihových variant (ASV), které byly shrnuty v předchozí kapitole, dávají vzniknout více než 20 tisícům ASV vznikajících z méně než 20 tisíc protein-kódujících genů. Tento odhad je odvozen z proteomických dat z databáze neXtProt (ver. 2015_06), díky které byly v lidském organismu identifikovány průměrně $1,1 \pm 0,1$ alternativní sestřihové varianty na jeden protein-kódující gen (z jednoho genu tedy vzniká 2,1 transkripční varianty; Ponomarenko et al., 2016).

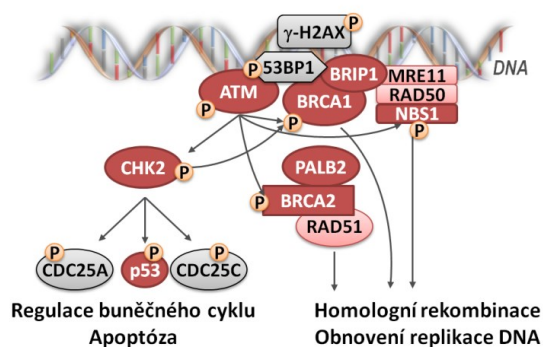
Naše laboratoř se v rámci svého onkogenetického zaměření zabývá geny predisponujícími ke vzniku dědičné formy karcinomu prsu a ovaria. Na základě našeho předchozího pilotního výzkumu identifikace ASV (Sevcik et al., 2010; Hojný, 2012; Hojny et al., 2017 - viz příloha 4) a funkční charakterizace vybraných ASV BRCA1 (Sevcik et al., 2012; Sevcik et al., 2013 – viz příloha 5) se tato práce dále zaměřuje na problematiku alternativního sestřihu *BRCA1* jako modelového genu, který je z pohledu ASV nejvíce probádán.

1.3.1 Predispoziční geny dědičné formy karcinomu prsu

Karcinom prsu je nejčastějším ženským onkologickým onemocněním v České republice se stoupající incidencí (134,4/100 tis. žen), a v posledních dvaceti letech mírně klesající mortalitou (35,5/100 tis. žen; ÚZIS ČR, data k roku 2016). Většina případů vykazuje formu sporadického onemocnění vznikajícího na základě kumulace somatických mutací v prsní tkáni postiženého jedince v průběhu života.

U zhruba 5 až 10 % ze všech případů karcinomu prsu je pozorována dědičná souvislost. Dědičná forma onemocnění vykazuje proti sporadické formě několik zásadních odlišností, zejména nástup onemocnění ve výrazně mladším věku, nádorové rekurence a multiplicity, často odlišný histologický typ či výskyt asociovaných nádorových onemocnění, jako je například karcinom ovaria. Postupem času byla na základě analýz genomové DNA v rodinách s dědičnou formou karcinomu prsu identifikována řada genů, jejichž patogenní mutace k tomuto subtypu nádorového onemocnění predisponují (Pohlreich et al., 2005; Lhota et al., 2016). Tyto geny můžeme rozdělit na geny významně zvyšující riziko vzniku tohoto onemocnění (relativní riziko > 5 ; *BRCA1* a *BRCA2*, *TP53*, *PTEN*) a geny se střední či nižší penetrancí (relativní riziko je mezi 2 a 5; *PALB2*, *ATM*, *CHEK2*, *BRIP1*, *RAD51C*, a další; Filippini a Vega, 2013).

Zajímavým faktem je přímá účast většiny jmenovaných genů na opravě nebo identifikaci / signalizaci vzniku dvouřetězcových zlomů DNA pomocí homologní rekombinace, či na regulaci těchto funkcí (Obr. 8; Bartkova et al., 2008).



Obrázek 8: Schéma proteinů, účastnících se homologní rekombinace na několika úrovních. Geny zároveň predisponující ke vzniku dědičné formy karcinomu prsu jsou vyznačeny červeně – ATM a MRN komplex – signalizace poškození; CHEK2 – přenos signálu; BRCA1, BRIP1, BRCA2, PALB2, RAD51 – vlastní oprava a její regulace; p53 – zastavení buněčného cyklu a aktivace apoptózy. Upraveno dle Bartkova et al., 2008.

Příčinné mutace v *BRCA1* jsou u zatížených rodin celorepublikově nejčastější. Kumulativní celoživotní riziko vzniku karcinomu prsu je u nosičů těchto mutací zároveň nejvyšší. U 80leté nosičky mutace bylo riziko vypočítáno na 72 % (u druhého nejčastěji mutovaného genu *BRCA2* je to 69 %; Kuchenbaecker et al., 2017). Dle databáze BIC (Breast Cancer Information Core - <https://research.nhgri.nih.gov/bic/>), která obsahuje 15311 záznamů alterací sekvence *BRCA1*, má 6133 (40 %) variant protein poškozující charakter (posuvné, nesmyslné, sestřihové a přestavby) a nachází se napříč celou kódující sekvencí (Anantha et al., 2017).

1.3.2 BRCA1

Gen *BRCA1* (Breast Cancer 1 Gene; OMIM 13705) byl definitivně identifikován v roce 1994, po intenzivním výzkumu oblasti predisponujícímu k dědičné formě karcinomu prsu na dlouhém raménku chromozomu 17 (17q21.31; Hall et al., 1990; Miki et al., 1994). Gen má velikost 81 188 párů bází (bp) a sestává se z 23 délkově výrazně heterogenních exonů (Obr. 9.)



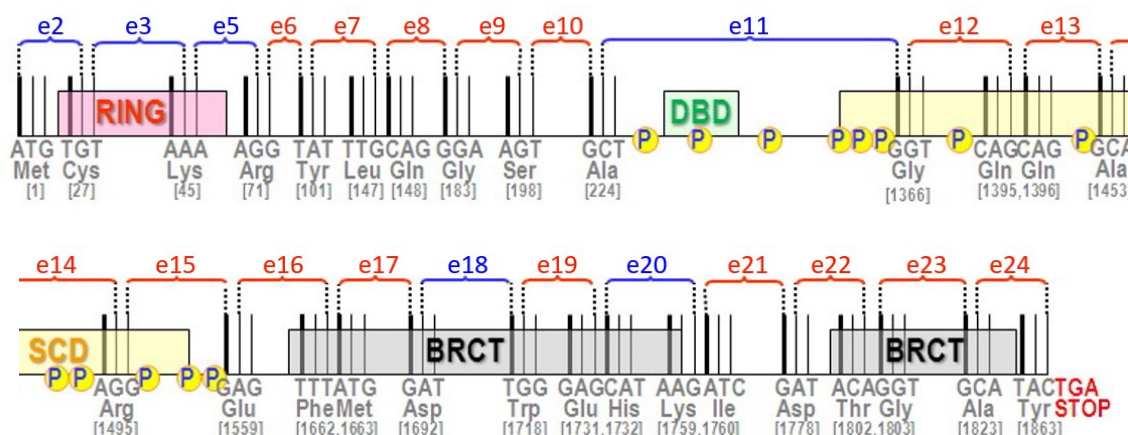
Obrázek 9: Složení a poměry délek kódujících exonů genu *BRCA1*. Nejdelším exonem je exon 11, který má velikost 3426 bp a zahrnuje tak více jak 60 % kódující sekvence. Nejmenší exon 19 obsahuje pouze 41 bp. Délka zobrazených exonů na obrázku poměrově odpovídá jejich velikosti (bp).

První exon je nekódující² a jedenáctý exon zaujímá více než 60 % kódující sekvence (Obr. 9; Smith et al., 1996).

1.3.2.1 Struktura transkriptu BRCA1

Kanonický transkript genu *BRCA1* má velikost 7,8 kb a obsahuje 23 exonů, z nichž 22 kóduje řadu funkčních domén napříč 1863 aminokyselinovým proteinem o velikosti 220 kD (Obr 10.), které zajišťují hlavní popisované funkce kanonické formy BRCA1.

Na 5' konci transkriptu je přes 3 exony lokalizována RING finger doména (Really Interesting New Gene) schopná vazby Zn^{2+} (aminokyseliny 24–65). Je to jedna z charakteristických E3-ubikvitinligázových domén (enzymatických domén schopných katalyzovat vazbu ubikvitinu na specifický cílový protein; Miki et al., 1994). Tato doména je zároveň schopna vazby se stejnou doménou jiných proteinů, což mnohonásobně zvyšuje katalytický potenciál vzniklého E3-ubikvitinligázového heterokomplexu (Lorick et al., 1999).



Obrázek 10: Schéma oblastí kódujících funkční domény v kanonickém transkriptu BRCA1.

Červeně – RING doména. **Zeleně** (DBD) – DNA vazebná doména. **Žlutě** (SCD) – doména bohatá na seriny. **Šedě** – BRCT domény. **Modře** jsou označeny exony, které nezpůsobují posun čtecího rámce po výpadku z transkriptu a **červeně** ty, které po výpadku posun způsobí. Obrázek je rozdělen na dvě části v exonu 14. Upraveno dle Sevcik et al., 2012; Hojný 2012.

² Dle původní nomenklatury chybí v transkriptu BRCA1 exon 4. Tento exon 4 je označován jako alternativní (původně byl chybně identifikován, jako součást kanonického transkriptu). Transkript dle původní, staré nomenklatury, obsahuje kódující exony 2-3 a poté 5-24 a kromě klasického exonu 13 případně i alternativní exon 13 (umístěn mezi exony 13 a 14).

V nové, aktualizované, nomenklatuře je pořadí exonů přepočítáno, „alternativní“ exon 4 vyřazen, a jako „nový“ exon 4 označen původní exon 5 atd. Pro změnu je však v nové nomenklatuře vynechán exon 13, kterým je označován původně alternativní exon 13. Dle nové nomenklatury, má BRCA1 kódující exony 2-12 a 14-24, přičemž exony 2 a 3 a exony 14-24 si u obou nomenklatur odpovídají. Exony 4 až 12 dle nové nomenklatury odpovídají číslování exonů 5 až 13 dle původní nomenklatury.

V této práci je používána původní nomenklatura exonů, která se v literatuře stále používá.

Na 3' konci transkriptu nalezneme oblasti kódující tandemové BRCT domény (BRCA1 C-terminal; AK 1642-1736 a 1756-1855; Williams et al., 2001). Tyto vazebné domény zajišťují specifickou vazbu na řadu konkrétních fosfoproteinových epitopů, díky které může BRCA1 vytvářet několik funkčně odlišných proteinových makrokomplexů, plnících různé úlohy v průběhu oprav dvouřetězcových zlomů DNA (Koonin et al., 1996; Bork et al., 1997).

V transkriptu BRCA1 jsou kódovány dva jaderné lokalizační signály (NLS; aminokyseliny 501-508 a 607-614), které jsou schopné interagovat s importinem α , avšak samy o sobě nezajišťují vstup BRCA1 do jádra (Chen et al., 1996; Thakur et al., 1997). Ten je zajištěn heterodimerní vazbou RING finger domén s proteinem BARD1, kdy je využíván silný NLS BARD1. Tato vazba zároveň kryje dva silné nukleární exportní signály (NES; aminokyseliny 22-30 a 81-99) v BRCA1 (Rodríguez a Henderson, 2000; Fabbro et al., 2002).

Mezi 3840. a 4572. kódující bázi transkriptu je lokalizován úsek, který dává vzniknout doméně bohaté na serinové zbytky. Serinová doména slouží jako aktivační doména proteinu BRCA. Fosforylací těchto zbytků řadou kináz (dominantně ATM a ATR, dále také CDK2, DNA-PKcs, CHK2, Akt1 a AURKA) je modulována konformace proteinu (zvláště pak C-koncových BRCT domén), která přímo ovlivňuje vazebnou kapacitu BRCA1 k řadě vazebných partnerů v průběhu oprav dvouřetězcových zlomů DNA homologní rekombinací (Ouchi et al., 2006).

Byla prokázána schopnost proteinu BRCA1 nespecifické vazby na jednořetězcové 300-500 bází dlouhé úseky DNA, která je zajišťována DNA-vazebnou doménou lokalizovanou mezi 452. a 1079. aminokyselinou. Tato schopnost je pravděpodobně využívána při opravách dvouřetězcových zlomů nebo při kolapsu replikační vidlice, kdy je velké množství molekul BRCA1 (či komplexů s BRCA1) schopno vázat volnou DNA a zabránit tak dalšímu poškození, a zároveň, společně s ostatními proteiny, modulovat opravu tohoto poškození (Paull et al., 2001).

Kromě DNA-vazebné domény a obou jaderných lokalizačních signálů jsou všechny ostatní domény kódovány vždy několika, po sobě jdoucími, exony. Změna v posloupnosti primárního transkriptu (ať už ve formě aberantního nebo alternativního sestřihu) by znamenala, zvláště pro konzervované RING a BRCT domény, významný zásah do popisovaných funkcí translatovaného proteinu.

1.3.2.2 Funkce proteinu BRCA1

Lokalizace a popsání genu *BRCA1* v roce 1994 zároveň odstartovalo intenzivní práci mnoha vědeckých skupin na objasnění funkce BRCA1 v lidském organismu. Postupem času bylo zjištěno, že hlavní funkcí tohoto jaderného fosfoproteinu a zároveň tumor-supresoru je udržování stability genomové DNA. Stabilita je udržována pomocí modulační opravy dvouřetězcových zlomů homologní rekombinací a opravy DNA při zhroucení replikační vidlice v průběhu replikace gDNA. Nepostradatelnost proteinu BRCA1 a jeho funkcí pro (lidský) organismus prokazuje experiment s myšími embryi *BRCA1*^{-/-}, které vykazovaly časnou letalitu (Hakem et al., 1996).

BRCA1 plní své funkce v organismu na několika úrovních pomocí proteinových multikomplexů, které se odlišují složením a konkrétní funkcí v rámci opravy dvouřetězcových zlomů DNA.

Zásadním vazebným partnerem BRCA1 je protein BARD1, který s BRCA1 interaguje na N konci pomocí RING finger domén obou proteinů, čímž je zajištěn transport BRCA1 do jádra (viz předchozí kapitola). Dále je touto vazbou mnohonásobně zvýšena E3-ubikvitinligázová aktivita heterokomplexu, která je využita při signalizaci během opravy zlomů (monoubikvitinylace histonu H2AX a proteinu CtIP; Dever et al., 2012; Mallery et al., 2002; Morris a Solomon, 2004). Komplex s BARD1 je také nazýván „core complex“ (volně přeloženo jako základní komplex), který je součástí dalších, „nadstavbových“ komplexů tvořených C koncem BRCA1.

C koncové multikomplexy jsou označovány podle klíčových interakčních partnerů BRCA1 – A (protein Abraxas), B (helikáza BRIP1) a C (CtIP) a dále je znám komplex označovaný jako BRCC (Huen et al., 2010).

- **BRCA1A** obsahuje proteiny Abraxas (ABRA1), RAP80, BRCC36, BRCC45 a NBA1 a jeho hlavní funkcí je lokalizovat BRCA1/BARD1 a BRCC36 do míst poškození DNA, kde tyto proteiny ubikvitinylují (BRCA1/BARD) / deubikvitinylují (BRCC36) své substráty, a tak regulují procesy opravy DNA (Wang et al., 2009; Dever et al., 2012).
- **BRCA1B** je složen z helikázy BRIP1 (známá také jako BACH1) a TOPBP1. Během S fáze buněčného cyklu se přímo účastní replikace DNA, případných oprav a signalizace poškození v jejím průběhu (Cantor et al., 2001, Huen et al., 2010).

- **BRCA1C** se skládá z proteinu CtIP a MRN komplexu (MRE11/RAD50/NBS1), který je v průběhu oprav pomocí homologní rekombinace zodpovědný za 5' – 3' resekci jednoho vlákna dvoušroubovice. Vznikající dlouhé jednořetězcové 3' přesahy jsou nezbytné pro správný průběh homologní rekombinace (Huen et al., 2010).
- **BRCC** komplex je tvořen proteiny RAD51, PALB2 a BRCA2. Jedná se o klíčový komplex průběhu homologní rekombinace, který je zodpovědný za vyhledávání homologních sekvencí na sesterské chromatidě. Ta poté slouží jako templát pro opravu poškozeného místa (Wong et al., 1997; Scully et al., 1997).

Z popisu je zřejmé, že BRCA1 je univerzální interakční partner proteinových komplexů účastnících se (nejen) homologní rekombinace na několika úrovních. Dá se uvažovat, že v závislosti na míře své fosforylace v doméně bohaté na seriny (od hlavních signálních a pomocných kináz) váže BRCA1 pomocí svých BRCT domén některého ze svých vazebných partnerů z komplexů A-C či BRCC. Míra či kombinace fosforylovaných serinů a dynamika fosforylace v průběhu oprav by se tak odrážela na dynamice vzniku BRCA1 komplexů, a tím na konkrétní fázi oprav dvouřetězcových zlomů DNA, či oprav spojených s poškozením během replikace DNA.

1.3.3 Alternativní sestřihové varianty BRCA1 a jejich význam

S ohledem na velikost kódující oblasti genu *BRCA1* a jeho členění do 22 exonů, byly (mutační) analýzy v raných experimentech obvykle prováděny na úrovni mRNA, což vyústilo v identifikaci a popis prvních ASV už dva roky po objevu BRCA1. Prvotně byly identifikovány a dále zkoumány varianty s výpadkem exonů 9 a 10 (BRCA1 Δ 9_10), varianta se zkráceným exonem 11 o jeho posledních 3309 bp (BRCA1 Δ 11q, někdy označovaná jen jako 11q) a kombinace těchto variant BRCA1 Δ 9_11,11q (Lu et al., 1996).

Ihned po jejich popsání se jejich funkčními analýzami začala zabývat řada skupin (Wilson et al., 1997; Wang et al., 1997; Chai et al., 2001) a první práce, popisující expresní profil zmíněných variant v různých buněčných liniích v průběhu buněčného cyklu, je z roku 2001 (Orban a Olah, 2001). V této práci je ukázána rozdílná míra exprese variant v buněčných liniích karcinomu prsu v porovnání s liniemi karcinomu ovaria a leukémie. Práce dále naznačuje možnost vlivu deregulovaného alt. sestřihu na tumorigenezi, i při absenci zárodečných mutací *BRCA1*.

V následující práci Orbana a Oláhové (2003) je poprvé publikován souhrn do té doby identifikovaných ASV BRCA1 se seznamem tkání, ve kterých byly detekovány (Tab. 1; Orban a Olah, 2003).

Tabulka 1: Souhrn alternativních sestřihových variant BRCA1 dle Orbana a Oláhové, 2003.

	Popis varianty	Tkáňový původ
Varianty zachovávající čtecí rámec	wt BRCA1	prs, ovarium, varle, brzlík a další
	- s exonem 1a	prs, ovarium, varle, brzlík
	- s exonem 1b	placenta
	$\Delta 2_{-10}$	prs, lymfocyty
	$\Delta 5$	prs, ovarium, lymfocyty
	$\Delta 6,5q$	prs, lymfocyty
	$\Delta 9_{-10}$	prs, ovarium, lymfocyty
	$\Delta 9_{-10,11q}$	prs, ovarium, lymfocyty
	$\Delta 9_{-11}$	prs, lymfocyty
	11q	prs, ovarium, lymfocyty
	$\Delta 11$	ovarium, štítná žláza
	$\Delta 14_{-17}$	prs, lymfocyty
	$\Delta 14_{-18}$	prs, lymfocyty
Varianty s posunem čtecího rámce	$\Delta 3$	prs, lymfocyty
	$\Delta 3,5q$	lymfocyty
	5q	prs, ovarium, lymfocyty
	$\Delta 6_{-7}$	lymfocyty
	$\Delta 9$	lymfocyty
	$\Delta 15_{-17}$	prs, lymfocyty
Identifikované varianty využívající volná sestřihová místa (LSS)	+ 3 bp 5' exon 8	prs
	+ 3 bp 5' exon 14	prs
	- 3 bp 5' exon 8	prs, lymfocyty
	- 3 bp 5' exon 12; 11q	prs
	- 6 bp 3' exon 1a	ledviny, plíce, další

V roce 2004 byla popsána zvláštní sestřihová varianta genu *BRCA1* – varianta BRCA1 IRIS (In-frame Reading of BRCA1 Intron Splice variant). Tato sestřihová varianta, shodně s kanonickou variantou, obsahuje kódující exony 2 až 11, po kterých však následuje 335 bází přilehlého intronu 11, ve kterém je přítomný STOP kodón a poly-A signál. Dále obsahuje varianta IRIS odlišnou část sekvence v exonu 1, který byl pojmenován jako exon 1c. Sekvence 5' části exonu 1c byla, dle práce, identifikována 24 Mb před samotným genem, což naznačuje využití jiného promotoru pro expresi transkriptu BRCA1-IRIS. Vzniklý proteinový produkt má stejnou proteinovou sekvenci, jako BRCA1, která je však „předčasně“ ukončena po exonu 11 inzercí 34 aminokyselin kódovaných přilehlou částí intronu 11 (ElShamy a Livingston 2004).

V rámci diagnostiky dědičné nádorové predispozice v letech 1997 až 2006 byly na úrovni mRNA v Laboratoři onkogenetiky Ústavu biochemie a experimentální onkologie 1. LF UK náhodně identifikovány alternativní sestřihové varianty BRCA1, včetně variant dosud nepopsaných (Tab. 2; Pohlreich et al., 2006).

Tabulka 2: Souhrn nalezených alternativních sestřihových variant BRCA1 v Laboratoři onkogenetiky na Ústavu biochemie a experimentální onkologie 1. LF UK identifikovaných v rámci diagnostiky dědičné nádorové predispozice na úrovni mRNA v letech 1997 až 2006. Barevně jsou zvýrazněny varianty do té doby nepopsané. Zeleně varianty zachovávající čtecí rámec. Červeně varianty s posunem čtecího rámce.

	Popis varianty	Tkáňový původ
Varianty zachovávající čtecí rámec	Δ3	lymfocyty periferní krve
	Δ3_5	lymfocyty periferní krve
	Δ5	lymfocyty periferní krve
	Δ9_10	lymfocyty periferní krve
	Δ9_11	lymfocyty periferní krve
	Δ13_19	lymfocyty periferní krve
	Δ14_15	lymfocyty periferní krve
	Δ17_19	lymfocyty periferní krve
Varianty s posunem čtecího rámce	Δ9	lymfocyty periferní krve
	Δ10	lymfocyty periferní krve
	Δ15	lymfocyty periferní krve
	Δ15_17	lymfocyty periferní krve
	Δ17	lymfocyty periferní krve
	Δ21	lymfocyty periferní krve

Z nálezů publikovaných do té doby bylo patrné, že z *BRCA1* se tvoří v průběhu exprese neobvykle velké spektrum různých transkriptů.

Rozsáhlá studie konsorcia ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles; <https://enigmaconsortium.org/>) z roku 2014 přinesla do studia alternativních sestřihových variant BRCA1 nové, zásadní informace. Kombinovanými a časově náročnými metodami fragmentační analýzy, klonováním a přímé sekvenace se podařilo identifikovat 63 nezávislých produktů alternativního sestřihu BRCA1, který zahrnuje 46 sekvenčně charakterizovaných variant a 17 variant odvozených kalkulací výsledků kapilární elektroforézy (Tab. 3; Colombo et al., 2014).

Tabulka 3: Souhrn nalezených variant dle Colombo et al., 2014. Varianty kazetového typu značí varianty s výpadkem nebo inzercí jednoho exonu / úseku intronu. **Varianty multikazetového typu** značí výpadek více exonů v řadě. **SDS / SAS** (splice site donor / acceptor shift) značí varianty s posunem sestřihového místa – tzv. volná sestřihová místa. **Intronizace** značí varianty s intronizovanou vnitřní částí exonu. **Terminální modifikace** značí varianty s předčasně ukončeným transkriptem. **Varianty smíšeného typu** značí varianty s kombinací dosud popisovaných subtypů.

Varianty kazetového typu	Varianty multikazetového typu	SDS / SAS	Varianty smíšeného typu
Δ2	Δ2,3	Δ2p	Δ1Aq,2
Δ3	Δ2_5	Δ8p	Δ1Aq_3
▼4	Δ2_10	Δ13p	Δ1Aq_5
Δ5	Δ8,9	Δ14p	Δ1Aq_10
Δ9	Δ8_10	Δ1Aq	Δ10_13p
Δ10	Δ9,10	▼1aA	Δ11_13p
Δ11	Δ9_11	Δ5q	Δ13_14p
Δ13	Δ9_12	Δ11q	▼13A,Δ14p
▼13A	Δ10,11	Intronizace	1B,Δ2
Δ14	Δ10_12		1B,Δ2,3
Δ15	Δ11,12	11Δ3110	1B,Δ2_5
Δ17	Δ14_15	11Δ3240	Δ2,3,▼4
Δ18	Δ14_17	Terminální modifikace	Δ1Aq,Δ2p
Δ20	Δ14_18		Δ1Aq_3,▼4
Δ21	Δ14_19	1B	
Δ22	Δ15_17	IRIS	
Δ23	Δ15_19		
	Δ21,22		
	Δ21_23		
	Δ22_23		

Jako hlavní tkáňové zdroje výzkumu byly uvedeny deriváty krevních buněk (imortalizované krevní buňky lymfoblastoidní buněčné linie a stimulované leukocyty periferní krve) a

dále leukocyty plné krve a monojaderné buňky periferní krve izolované pomocí ficolu. V projektu byl také analyzován jeden vzorek zdravé prsní tkáně, ve kterém bylo identifikováno 43 z 63 celkově nalezených variant.

Romero a kolektiv pak ve své práci z roku 2015 (Romero et al., 2015) doplnili předcházející studii analýzou nádorové prsní tkáně od 70 pacientů, analýzou 4 vzorků zdravé prsní tkáně a analýzou 72 zdravých kontrol na vzorcích periferní krve. V této práci, ve které byl použitý stejný metodický přístup jako u Colombo et al., 2014, nebyla nalezena žádná nová varianta.

Funkční analýzy provedeny v Laboratoři onkogenetiky 1. LF UK na vybraných ASV s výpadkem exonů kódujících SCD ($\Delta 14-15$) a výpadkem exonů kódujících první BRCT domény ($\Delta 17-19$), které byly náhodně identifikovány u pacientů s predispozicí ke karcinomu prsu, prokázaly velmi pomalou odpověď na poškození DNA a neschopnost správného sestavení DNA opravného komplexu, tedy významně poškozenou funkčnost výsledného proteinu BRCA1 (Sevcik et al., 2012; Sevcik et al., 2013).

Vzhledem k identifikaci řady variant, které postrádají exony nebo části exonů kódujících důležité funkční domény ($\Delta 3$; $\Delta 3-5$; $\Delta 11$; $11q$; $\Delta 13$; $\Delta 14$; ...), vyplynula nutnost objasnit kvantitativní složení ASV BRCA1 v relevantních zdravých (i patologických) tkáních člověka, kterými jsou leukocyty periferní krve (tkáň, ze které je nejčastěji izolována genetická informace) a prsní a přilehlá tuková tkáň (která je postižena případným karcinomem prsu).

2 CÍLE PRÁCE

Jako modelový gen pro analýzu alternativních sestřihových variant, který predisponuje ke karcinomu prsu, byl vybrán gen *BRCA1*. Důvodem byla existence a precizní popsání řady jeho ASV, dlouhodobý výzkum tohoto genu v naší laboratoři, dostatečná komplexita (obsahuje 22 kódujících exonů) a v neposlední řadě velmi nízká exprese mRNA ve zdravých tkáních. RPKM se ve většině vyšetřovaných tkání pohybuje okolo 1 (Reads Per Kilobase Million – normalizovaná hodnota množství čtení na kilobázi transkriptu vztaženou k milionu čtení; Fagerberg et al., 2014). Z těchto důvodů je analýza ASV *BRCA1* obvyklými metodami nelehká.

Cíle práce zahrnovaly:

- a) vytvořit robustní systém s vysokou senzitivitou a reprodukovatelností pro rychlou a komplexní identifikaci přítomných alternativních sestřihových variant mRNA *BRCA1*, který by byl univerzální a aplikovatelný na jakýkoliv jiný gen našeho zájmu
- b) identifikovat tímto systémem alternativní sestřihové varianty mRNA *BRCA1* a pomocí RT-qPCR s primery, nasedajícími do specifických exon-exonových spojení, kvantifikovat vybrané sestřihové varianty a vytvořit tak komplexní kvalitativní i kvantitativní katalog mRNA variant *BRCA1*, vyskytujících se v leukocytech periferní krve, nenádorové prsní tkáni a nenádorové perimamární tukové tkáni
- c) vytvořit a vyhodnotit expresní profil sestřihových variant mRNA *BRCA1* ve vyšetřovaných tkáních u pacientek s karcinomem prsu a zdravých kontrol.

3 MATERIÁL A METODY

3.1 Strategie řešení

V přístupech pro kvalitativní a kvantitativní analýzu alternativních sestřihových variant byla navržena metodika využívající kombinaci jednoduchých, rutinně prováděných, dílčích technik tak, aby měl získaný přístup vysokou senzitivitu, byl jednoduše opakovatelný, a navíc cenově přijatelný. Jedná se o:

- a) Multiplexní PCR (mPCR) amplifikaci všech teoreticky možných exon-exonových spojení všech transkriptů požadovaného genu (model *BRCA1*) na úrovni komplementární DNA (cDNA) syntetizované z mRNA pomocí náhodných hexanukleotidů.
- b) Ekvimolární smísení mPCR produktů a velikostní selekce (obohacení) oblasti s krátkými amplikony našeho zájmu.
- c) Sekvenování na základě NGS metod s hlubokým pokrytím pro rychlou analýzu vzniklých amplikonů, které vykazuje vysokou citlivost a relativně snadnou přípravu vzorků. Analýza sekvenačních dat pomocí mapování získaných amplikonů na uměle vytvořený referenční soubor obsahující všechny možné kombinace sestřihových variant společně s variantami dosud identifikovanými.
- d) Zhodnocení míry exprese vybraných sestřihových variant zachovávajících čtecí rámec, u kterých předpokládáme vznik proteinového produktu, pomocí kvantitativní PCR v reálném čase s využitím primerů, které specificky nasedají do unikátních exon-exonových spojení.

3.2 Charakteristika vyšetřovaného souboru

3.2.1 Biologické vzorky od pacientek s karcinomem prsu a kontrol

Do projektu bylo zařazeno celkem 130 žen operovaných pro karcinom prsu na Gynekologicko-porodnické klinice 1. LF UK a VFN a 86 žen bez nádoru operovaných na Klinice plastické chirurgie 1. LF UK a FNB. Všechny pacientky byly bělošky české národnosti původem z oblasti Čech a vyjádřily souhlas se zařazením do studie podpisem informovaného souhlasu schváleného etickou komisí 1. LF UK a VFN. Od všech účastnic byla získána relevantní data z osobní a rodinné anamnézy (Tab. 4).

Tabulka 4: Přehled souborů pacientek a zdravých kontrol s mutací a bez mutace v *BRCA1*

	S mutací v <i>BRCA1</i>	<i>BRCA1</i> wt	Celkem
Ženy bez karcinomu prsu [počet]	10	76	86
- průměrný věk při odebrání vzorku (rozsah) ± SD; [roky]	42,2 (26-59) ± 9,7	41,1 (18-66) ± 13,1	41,2 (18-66) ± 12,7
Pacientky s karcinomem prsu [počet]	7	123	130
- průměrný věk při odebrání vzorku (rozsah) ± SD; [roky]	35,9 (28-41) ± 4,9	58,7 (31-83) ± 11,1	57,5 (28-83) ± 9,3
Celkem [počet]	17	199	216
- průměrný věk při odebrání vzorku (rozsah) ± SD; [roky]	39,6 (26-59) ± 8,4	52,0 (18-83) ± 14,7	51,0 (18-83) ± 14,6

3.2.2 Buněčné linie

Z důvodu omezeného množství celkové RNA, izolované ze tkání vyšetřovaných žen, proběhly optimalizace dílčích kroků (syntéza cDNA s náhodnými hexanukleotidy, multiplexní PCR, velikostní selekce, přípravy sekvenačních knihoven a sekvenování nové generace, RT-qPCR reakce, preamplifikace vzorků pro RT-qPCR) s využitím celkové RNA izolované z buněčných linií MCF7 (<http://www.mcf7.com>; Soule et al., 1973), EM-G3 (Brozova et al., 2007), HeLa (<http://www.hela-transfection.com>; Scherer et al., 1953) a MDA-MB-231 (Cailleau et al., 1974).

3.3 Odběr, izolace a zpracování biologického materiálu

3.3.1 Odběr vzorků

Tkáňové vzorky byly odebírány během operací karcinomu prsu nebo během profylaktických mastektomií provedených na asymptomatických nosičkách s *BRCA1* mutací. Kontrolní tkáňové vzorky byly odebrány během kosmetických operací. Pokud to bylo možné, byly vzorky odebírány z obou prsních tkání. Peroperačně získané vzorky byly neprodleně (do 5 minut) vloženy do zkumavek s roztokem RNAlater (Invitrogen) v poměru doporučeném výrobcem (cca 100 mg tkáně do 1 ml prezervačního média). Po 24h inkubaci při teplotě 4 °C byly vzorky dále skladovány v -80 °C do doby izolace RNA.

Od každé ženy byla získána periferní nesrážlivá krev, která byla odebrána do zkumavek PAXGene Blood RNA System (PreAnalytiX). Zkumavky s krví byly inkubovány 2 hod. při pokojové teplotě a následně uskladněny při -20 °C do doby izolace RNA.

3.3.2 Izolace celkové RNA z buněčných linií

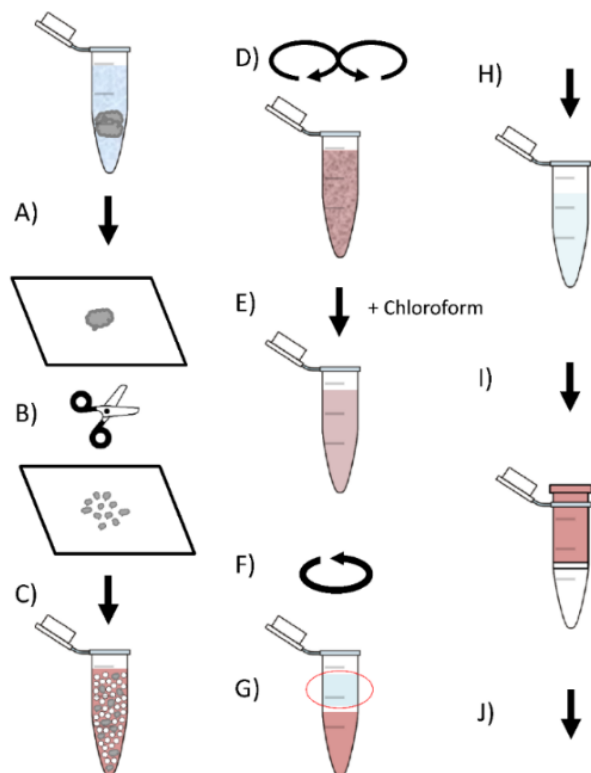
Celková RNA pro optimalizační pokusy byla izolovaná z buněčných linií s využitím izolačního kitu RNeasy Mini (Qiagen) dle protokolu výrobce pro izolaci celkové RNA z buněčných linií (Purification of Total RNA from Animal Cells Using Spin Technology). Na závěr izolace byla RNA eluována do 100 µl sterilní ddH₂O bez přítomnosti RNáz. Izolované vzorky RNA byly skladovány při -80 °C.

3.3.3 Izolace celkové RNA z periferní krve

Po rozmražení prezervačních PAXGene zkumavek s periferní žilní krví o objemu 2,5 ml byla celková RNA z bílých krvinek izolována kitem PAXGene Blood RNA Kit (PreAnalytiX) dle protokolu výrobce (včetně odstranění genomové DNA na kolonce pomocí DNázy I). Na závěr izolace byla RNA eluována do 100 µl sterilní ddH₂O bez přítomnosti RNáz. Izolované vzorky RNA byly dále do doby analýzy skladovány při -80 °C.

3.3.4 Izolace celkové RNA ze solidních tkání

Izolace RNA ze solidních tkáňových vzorků byla provedena dle schématu (Obr. 11):



Obrázek 11: Schéma izolace RNA ze solidních tkání.

- A) Odebrání 30-50 mg tkáně z RNAlateru na sterilní pracovní sklo.
B) Rozstříhání tkáně sterilními nůžkami na drobné kousky.
C) Vložení drobných kousků do zkumavky s homogenizačními kuličkami a QIAzol Lysis roztokem.
D) Homogenizace 3x15 vteřin.
E) Odpipetování homogenátu, krátká inkubace při pokojové teplotě a smísení s chloroformem, protřepání.
F) Centrifugace 15 min. při 12000 ot./min. a teplotě -2 °C.
G) Opatrné odebrání horní vodné fáze.
H) Připipetování odebrané vodné fáze do zkumavky se 70% etanolem, třepání.
I) Přepipetování směsi do izolační / purifikační Qiagen RNA kolonky.
J) Přechystání vzorku na kolonci dle standardního návodu kitu RNeasy Lipid Tissue Mini.

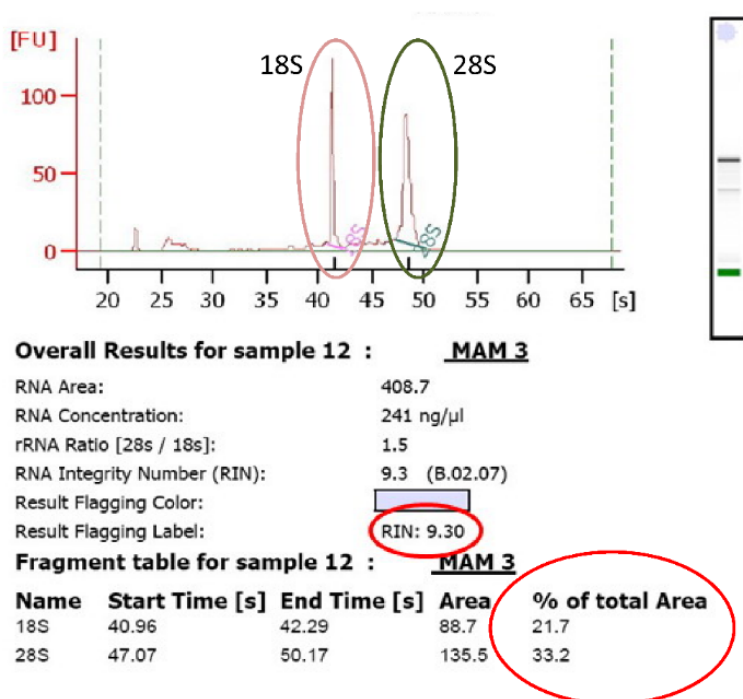
Ze skladovaných tkáňových vzorků byla po rozmražení z RNAlateru část vyjmuta (30–50 mg). Větší kusy tkáně byly sterilními nůžkami na sterilním pracovním skle rozstříhány na drobné kousky a vloženy do homogenizačních zkumavek (MagNA Lyser Green Beads; Roche) s 1 ml roztoku QIAzol Lysis Reagent (Qiagen). Pro vlastní izolaci byla tkáň homogenizována po dobu 3 x 15 vteřin při 7000 ot./min. na přístroji MagNA Lyser Instrument (Roche). Mezi jednotlivými homogenizačními běhy byl vzorek inkubován na ledu po dobu 1 minuty. Vzorek byl poté krátce centrifugován a homogenát byl odebrán do nové zkumavky. Dále byl vzorek 5 minut inkubován při pokojové teplotě a následně bylo ke vzorku přidáno 200 µl chloroformu. Zkumavka byla protřepána a centrifugována 15 min, při 12000 ot./min. a teplotě -2 °C. Poté byla odebrána horní vodná fáze (cca 400 µl). K vodné fázi byl přidán stejný objem 70% etanolu. Roztok byl 15 s třepán na třepačce a následně byl dále zpracován kitem RNeasy Lipid Tissue Mini (Qiagen) dle přiloženého protokolu (Protocol for Isolation of Total RNA from Animal Tissues) od kroku číslo 7. RNA byla eluována do 40 µl sterilní ddH₂O bez přítomnosti RNáz. Izolované vzorky RNA byly dále do doby analýzy skladovány při -80 °C. (Hojný J., 2012).

3.3.5 Kontrola kvality izolované RNA

Pro potřeby kvantitativní RT-PCR bylo nezbytné izolovat dostatečné množství RNA s co nejnižší mírou degradace. Míru degradace ovlivňuje každý krok práce s RNA od odběru tkáně až po syntézu cDNA. Čím je vstupní RNA fragmentovanější, tím větší je chyba při hodnocení míry exprese. RT-qPCR analýza z vysoce degradované RNA (cDNA) neodráží skutečné množství RNA ve vzorku, protože se vyšetřované amplikony neamplifikují stejnou měrou jako u intaktní RNA (cDNA).

Integrita celkové RNA byla charakterizována kapilární gelovou elektroforézou na přístroji Agilent 2100 Bioanalyzer (Agilent Technologies) za použití kitu Agilent RNA 6000 Nano Kit (Agilent Technologies) dle protokolu výrobce. Výsledkem analýzy bylo RIN (RNA integrity number), číslo v rozmezí 0-10, kde 10 značí celkovou RNA nejvyšší kvality. RIN je vygenerováno na základě algoritmů porovnávajících množství 18S a 28S rRNA oproti celkovému množství veškeré RNA (Obr. 12).

Průměrné RIN celkové RNA izolované z krve bylo 7,4 (min = 2,5; max = 9,1; standardní odchylka = $\pm 1,1$). Průměrné RIN celkové RNA izolované ze solidních tkání bylo 6,9 (min = 2,2; max = 9,4; standardní odchylka = $\pm 0,97$). Vzorky RNA s RIN <5 byly z další analýzy vyloučeny.



Obrázek 12: Princip hodnocení integrity celkové RNA pomocí Agilent 2100 Bioanalyzer. Růžovou a zelenou elipsou jsou na výsledném elektroforegramu vyznačeny 18S a 28S podjednotky rRNA, červenými elipsami je vyznačen poměr ploch jednotlivých podjednotek oproti celkové RNA a výsledné přiřazené RIN číslo.

Množství izolované celkové RNA bylo kvantifikováno spektrofotometricky na přístroji NanoDrop 1000 (Thermo Scientific). Po kvantifikaci a charakterizaci kvality celkové RNA ihned následovala syntéza cDNA.

3.3.6 Syntéza cDNA

Před samotnou syntézou cDNA byly z RNA mamární tkáně a přilehlé tukové tkáně odstraněny případné zbytky DNA pomocí DNázy I (1 U/10 μ l RNA; Thermo Fisher Scientific) po dobu 30 minut při 37 °C. Inaktivace DNázy byla provedena 10minutovou inkubací při 65 °C za přítomnosti 5mM EDTA. RNA izolovaná ze vzorků krve byla ošetřena DNázou I v průběhu izolace dle protokolu výrobce.

Syntéza komplementární DNA byla provedena pomocí SuperScript III reverzní transkriptázy (SSIII; rekombinantní virová RNA-dependentní DNA polymeráza; SuperScript III First-Strand Synthesis System; Invitrogen, Thermo Fisher Scientific) a náhodných hexamerů (Roche) dle doporučení výrobce v objemu 20 μ l a následujícího protokolu:

Reagencie	Objem [μ l]
RNA ošetřená DNázou	13,2
Hexanukleotidy (30 pmol/ μ l)	0,5
50x dNTPs	0,8
65 °C (10 min); poté zchlazení na ledu	
DTT	1,0
5x pufr	4,0
SSIII	0,5
25 °C (5 min); 50 °C (50 min); 70 °C (15 min); 4 °C (∞)	

Celkem 1,5 μ g celkové RNA bylo vloženo do syntézy cDNA, která sloužila pro mPCR/NGS identifikaci ASV. 800 ng celkové RNA bylo vloženo do syntézy cDNA, která byla použita pro kvantitativní analýzu vybraných ASV pomocí RT-qPCR.

3.3.7 PCR kontrola reverzní transkripce syntetizované cDNA pro identifikaci ASV

Před krokem identifikace ASV BRCA1 pomocí NGS, kde se pracuje se směsnými mixy cDNA, bylo nutné ověřit úspěšnost syntézy cDNA všech dílčích vzorků. Ověření proběhlo pomocí PCR amplifikace s mPCR primerovými páry lokalizovanými do exonů 2 a 8 a dále do exonů 15 a 17 (B1 HTS e2F a B1 HTS e8R – 425 bp; B1 HTS e15F a B1 HTS e17R – 374 bp). PCR reakce byla připravena s pomocí FastStart Taq DNA kitu (Roche) dle základního doporučení výrobce v objemu 10 μ l se vstupním objemem 1 μ l cDNA a 600nM finální koncentrací každého primeru za použití PCR programu:

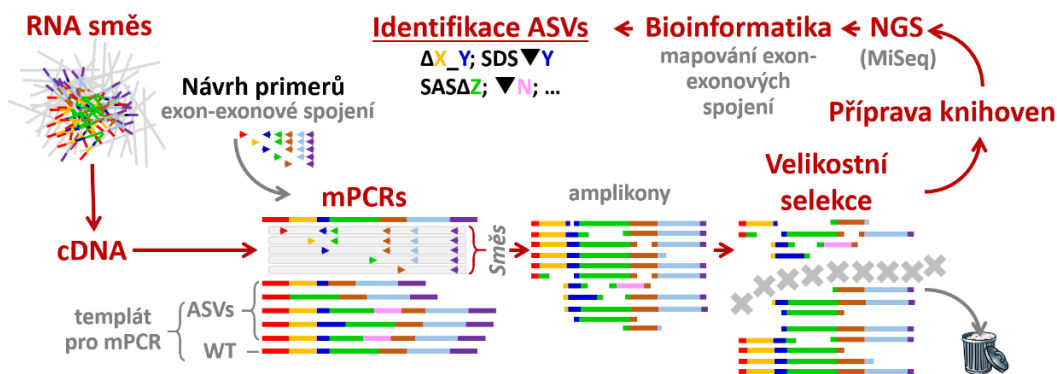
94 °C	5 min	} 35 x
94 °C	10 s	
62 °C	20 s	
72 °C	2 min	
72 °C	7 min	
4 °C	∞	

PCR produkt o objemu 1 μ l byl vizualizován pomocí 1,5% gelové elektroforézy (viz kapitola 4.2.1 *Kontrola správného průběhu syntézy cDNA pro identifikaci ASV*).

3.4 NGS identifikace alternativních sestřihových variant BRCA1

Většina prací, které se doposud věnovaly identifikaci alternativních sestřihových variant BRCA1 (viz kapitola 1.3.3 *Alternativní sestřihové varianty BRCA1 a jejich význam*), byly obvykle založeny na PCR amplifikaci celého transkriptu (5' UTR → 3' UTR) tak, aby nebyly opomenuty žádné alternativní sestřihové varianty (za předpokladu, že všechny ASV obsahují oba UTR). Bylo však prokázáno, že PCR amplifikace celého dlouhého transkriptu (v případě *BRCA1* až s 5,7 kilobází dlouhými mRNA) je velmi náročná na kvalitu izolovaného materiálu (u dlouhých mRNA dochází k degradaci a reverzní transkripce pomocí oligo-dT primeru má nízkou účinnost). V průběhu PCR pak dochází k preferenční amplifikaci extrémně krátkých ampliconů (pravděpodobně chybně sestřižených, odpadních mRNA), citlivost analýzy je u nízko exprimovaných genů, jako je *BRCA1*, pomocí klasického Sangerova sekvenování nízká a získané výsledky jsou obtížně reprodukovatelné (Hojný, 2012). Analýza celých transkriptů s vysokou citlivostí pro identifikaci i nízko-exprimovaných alternativních mRNA a dobrou reprodukovatelností je z důvodu vysokých nároků na množství odebíraného genetického materiálu, jeho uchovávání, zpracování v kombinaci s dostatečně velkým souborem vzorků v našich podmínkách, velmi obtížná.

Aby byly výše zmíněné obtíže eliminovány, byla navržena identifikace ASV pomocí multiplexní PCR všech exon-exonových spojení, obohacení produktů mPCR reakcí o krátké amplicony našeho zájmu pomocí velikostní selekce na SPRI (Solid Phase Reversible Immobilisation) magnetických kuličkách, příprava sekvenačních knihoven z těchto vzorků a finální sekvenování pomocí NGS s vysokým pokrytím a příslušnou biostatistickou analýzou sekvenačních dat (Obr. 13)



Obrázek 13: Schéma identifikace alternativních sestřihových variant BRCA1. Červenými šipkami je naznačen postup zpracování genetického materiálu od RNA až po identifikaci jednotlivých variant.

Výhodou tohoto přístupu amplifikace je relativně nízký nárok na intaktní RNA transkript. Relativně krátké cDNA fragmenty, vytvořené pomocí náhodných hexanukleotidů, pokrývají uniformně celou délku mRNA transkriptů, ze kterých jsou syntetizovány. Dále je minimalizována preferenční PCR amplifikace ampliconů, protože délka navržených PCR fragmentů exon-exonových spojení je krátká a velmi podobná. Naopak nevýhodou tohoto přístupu je ztráta informace o kombinaci nalezených sestřihových variant a omezená identifikace intronových inzercí (exonizace) a variant vznikajících na základě kryptických sestřihových míst. Tyto nevýhody jsou však nutným ústupkem rychlosti, citlivosti, robustnosti a finanční nenáročnosti navrženého postupu.

3.4.1 Multiplexní PCR

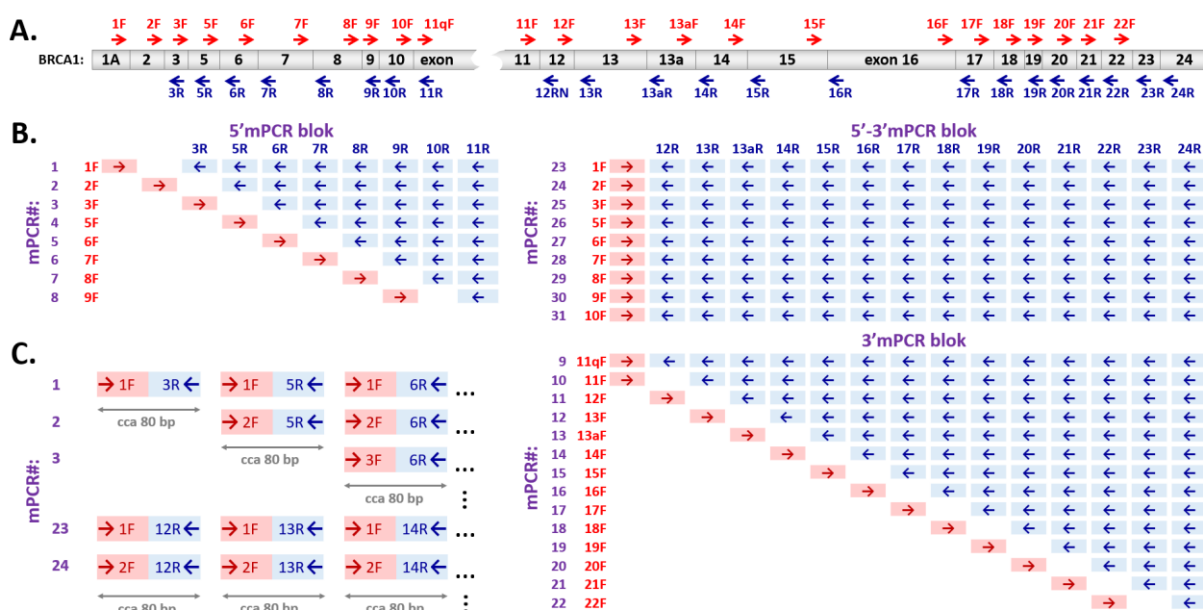
3.4.1.1 Návrh multiplexní PCR (mPCR)

Primery pro mPCR byly navrženy specificky do oblastí exon-exonových přechodů tak, aby dopředný (forward; FWD) primer byl umístěn do oblasti ≈ 40 párů bází před 3' konec exonu a zpětný primer (reverse; REV) byl umístěn ≈ 40 párů bází za 5' začátek dalšího exonu (příloha 1; Obr. 14 A). Takto byl získán amplicon o velikosti ≈ 80 párů bází bez ohledu na konkrétní exon-exonový přechod (Obr. 14 C). Velikost cca 80 páru bází byla navržena z důvodu přečtení celého fragmentu pomocí uvažovaného NGS sekvenování (Illumina) o velikosti 75 bází z každé strany fragmentu a z důvodu „rezervy“ pro případné odchylky od kanonických sestřihových míst (viz kapitola 1.2.2.5. *Alternativní sestřih na základě volných sestřihových míst*).

Varianta BRCA1 mRNA (NM_007294), označovaná v této práci jako wt či referenční, obsahuje 23 kanonických exonů. Dále je známa varianta NM_007300, která obsahuje 24 exonů, kdy je navíc přítomen exon vyskytující se mezi exony 13 a 14 předchozího transkriptu. Tento exon je v práci označován jako alternativní exon 13 (13A / 13alt). Také jsou popsány 2 případy alternativně zkrácených exonů (označovány 5q – NM_007297 a 11q – NM_007299; viz kapitola 1.3.3. *Alternativní sestřihové varianty BRCA1 a jejich význam*). Pokud všechny tyto varianty zkombinujeme, dostaneme 253 jednotlivých teoreticky možných exon-exonových spojení (22 kanonických spojení a 231 teoreticky vzniklých na základě alternativního sestřihu). Pro všechny jednotlivé PCR byly na arteficiálním templátu (plazmid obsahující referenční sekvenci *BRCA1*) a na testovacích RNA z buněčných linií, provedeny optimalizační reakce, díky kterým bylo možné ověřit, zda spolu jednotlivé navržené primery navzájem fungují za

stejných podmínek v PCR (viz kapitola 4.2.2 *Optimalizace primerů pro multiplexní PCR reakce*). Amplifikovat spojení po jednotlivých PCR reakcích je neefektivní, zejména s ohledem na omezené množství dostupného biologického materiálu, proto bylo přistoupeno k multiplexní PCR. Byla navržena řada mPCR reakcí, kde v jedné reakci byl přítomen vždy jeden FWD primer a sada REV primerů z následujících exonů (Obr. 14 B). Takto se počet reakcí sníží na počet exonů v transkriptu mínus jednu reakci.

Výstupem mPCR reakcí bez dodatečné úpravy by byl majoritní PCR produkt odpovídající běžně se vyskytujícímu kanonickému exon-exonovému spojení. Amplifikace nežádoucích kanonických přechodů proto byla omezena přeskočením REV primeru nasedajícího do následujícího kanonického exonu po exonu s přítomným FWD primerem (Obr. 14 B). Výsledné krátké ≈ 80 bp amplikony obsahují pouze exon-exonová spojení, která vznikla na základě alternativního sestřihu (Obr. 14 C).

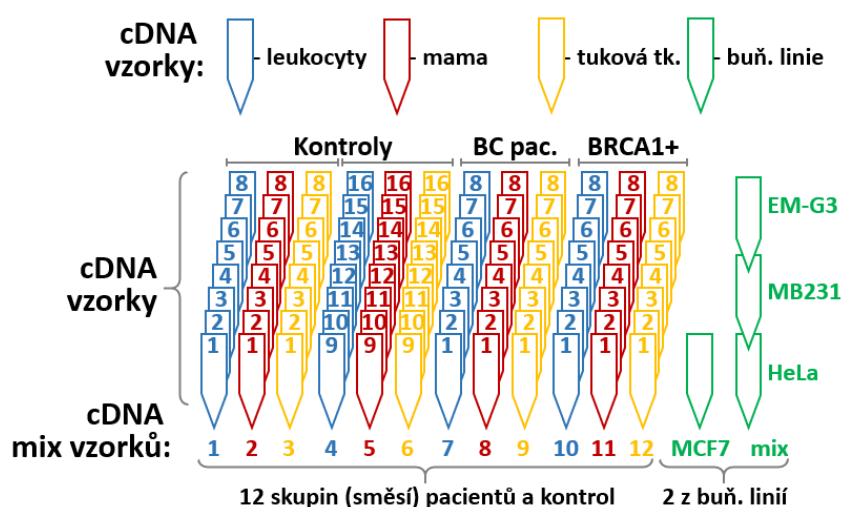


Obrázek 14: Schéma návrhu primerů pro mPCR. A) Přehled všech navržených primerů a jejich umístění napříč kanonickým BRCA1 transkriptem. B) Kombinace primerů v jednotlivých mPCR směsích. C) Příklad teoretických amplikonů vzniklých na základě amplifikace alternativně sestřiženého templátu.

S ohledem na extrémně dlouhý exon 11 (3426 bází) uprostřed wt transkriptu BRCA1, byl při navrhování mPCR reakcí celý transkript BRCA1 mRNA rozdělen na 3 pomyslné části tak, aby byly omezeny možné nežádoucí primerové interakce. Celkem bylo rozdělením BRCA1 transkriptu navrženo 31 jednotlivých mPCR reakcí (Obr. 14 B).

3.4.1.2 Vlastní mPCR reakce

Pro vlastní multiplexní PCR byly vybrány cDNA vzorky vzniklé syntézou z mRNA s nejvyšší kvalitou RIN (průměr 7,42; v rozmezí 4,3 – 8,9) od 16 žen ze skupiny zdravých kontrol (8 v každé ze dvou skupin), 8 žen ze skupiny pacientek s karcinomem prsu a 8 žen ze skupiny pacientek s BRCA1 mutací. Celkem tedy 4 skupiny po 8 pacientech. Od všech pacientek byly k dispozici 3 cDNA vzorky – původem z leukocytů periferní krve, prsní a přilehlé tukové tkáně. Dohromady bylo získáno 12 pracovních skupin (3 tkáně x 4 skupiny pacientů; Obr. 15).



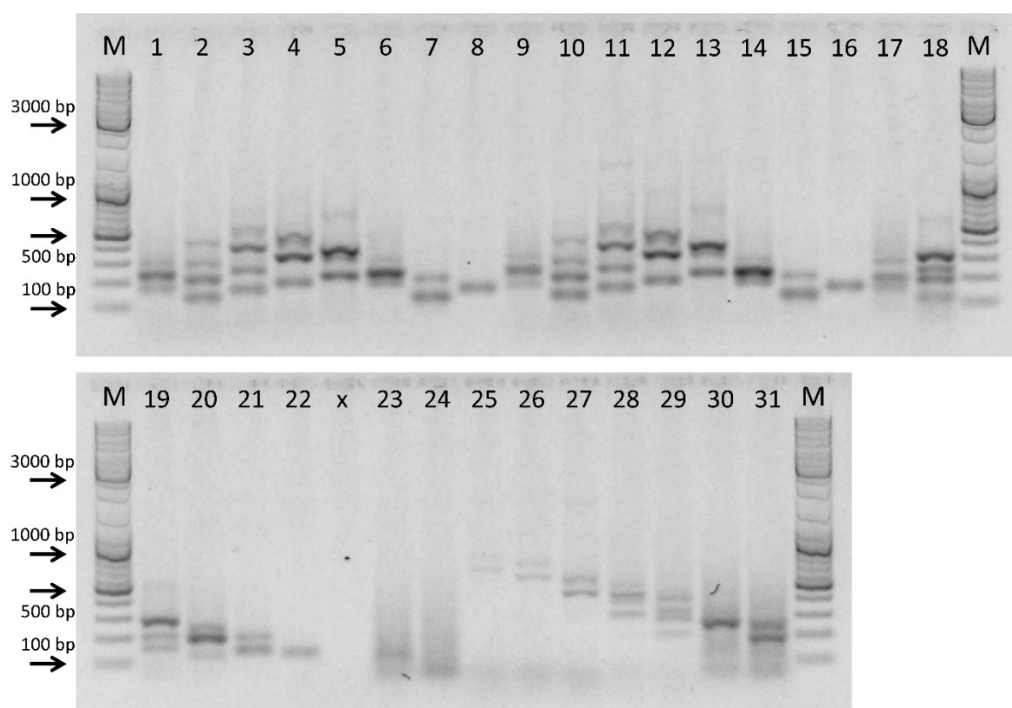
Obrázek 15: Schéma přípravy směsných vzorků pro mPCR reakce. Z každé skupiny žen (2x zdravé **kontroly**; 1x pacientky s karcinomem prsu – **BC pac.**; 1x nosičky BRCA mutace – **BRCA1+**) bylo vybráno 8 žen s nejvyšší kvalitou RNA. Osm vzorků stejného tkáňového původu bylo ekvimolárně smíšeno dohromady, což dalo vzniknout 12 směsným vzorkům. Další dva vzorky byly vytvořeny v rámci optimalizace metody z RNA buněčné linie MCF7 a směsi buněčných linií EM-G3, MB231 a HeLa.

Každý jednotlivý cDNA vzorek (18 µl) byl smíchán dohromady s ostatními vzorky stejného tkáňového původu ve skupině tak, aby byl poměr vstupního množství jednotlivých vzorků cDNA shodný. Ve vzniklých 12 směsích tak byly zachovány varianty opravdu alternativní, které se vyskytují napříč vybranými vzorky, a nařaděny varianty aberantní, vyskytující se individuálně na základě alterací genomu jedince.

Každá směs cDNA sloužila jako templát pro sadu 31 mPCR reakcí (4 µl cDNA na reakci, ekvivalent 300 ng RNA), která byla provedena v objemu 40 µl pomocí FastStart Taq DNA Polymerase kitu (Roche) dle základního doporučení výrobce s finální koncentrací 225 nM FWD primeru a mixu 75nM REV primerů a dle PCR programu:

94 °C	1 min	} 35x
94 °C	10 s	
62 °C	20 s	
72 °C	15 s	
72 °C	7 min	
15 °C	∞	

Správný průběh jednotlivých mPCR byl zkontrolován pomocí 1,5% gelové agarózy produktů (2,5 µl) všech jednotlivých reakcí (Obr. 16).

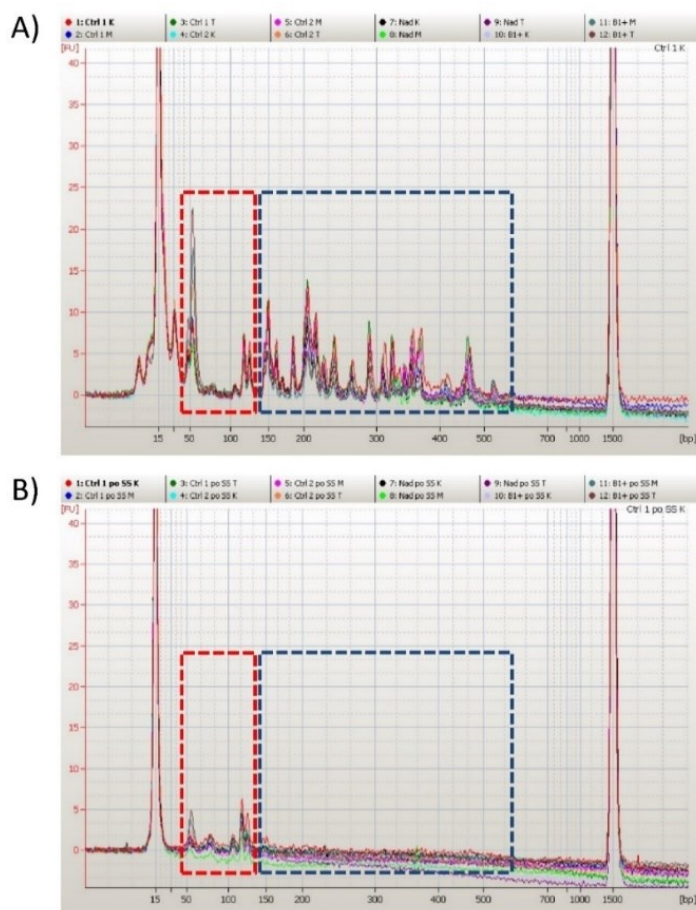


Obrázek 16: Příklad kontrolní elektroforézy všech jednotlivých mPCR reakcí. Čísla 1-31 značí jednotlivé mPCR směsné reakce. M – velikostní marker. x – vynechaný pruh. Z elektroforézy je patrné, že každá mPCR tvořila produkt (včetně reakce 23 a 24, pro které je slabý neostrý produkt v oblasti odpovídající 50-150 bp typický). Většina detekovatelných pruhů odpovídá produktům získaným z kanonického wt templátu, slabé produkty alternativních sestřihových variant o velikosti přibližně 80 bází jsou z důvodu nízké koncentrace na klasické gelové elektroforéze obtížně detekovatelné (výjimkou jsou reakční směsi 23-27).

Produkty (35 µl) sady mPCR z každé z 12 skupin byly zpětně smíchány dohromady a poté charakterizovány pomocí kapilární gelové elektroforézy na čipu (Agilent 2100 Bioanalyzer) za využití kitu DNA 1000 (Agilent Technologies; Obr. 17 A) dle doporučení výrobce.

3.4.2 Velikostní selekce

Pro obohacení vzorku o krátké amplikony (s alternativními exon-exonovými přechody) na úkor dlouhých amplikonů (s pravděpodobně kanonickými přechody) bylo přistoupeno k velikostní selekci za použití Agencourt AMPure XP Reagent (Beckmann Coulter) dle instrukcí výrobce. Tento systém využívá vlastnosti vazby DNA na porézní magnetické kuličky za přítomnosti pufru obsahujícího polyetylglykol. Čím je vyšší koncentrace pufru, tím menší DNA fragmenty se váží na magnetické partikule. Pomocí tohoto systému byla provedena 2 přečištění v řadě. Během prvního přečištění s koncentrací pufru 1,8x jsou dlouhé fragmenty navázány na kuličky a supernatant obsahuje krátké amplikony našeho zájmu. Po přepipetování supernatantu je společně s kuličkami odstraněna většina fragmentů > 150 bp. V následném přečištění původního supernatantu s koncentrací pufru 2,5x jsou odstraněny fragmenty < 50 bp, amplikony našeho zájmu jsou vychytány magnetickými partikulami, které jsou následně dle protokolu výrobce přečištěny. Výsledek velikostní selekce byl zkontrolován kapilární gelovou elektroforézou na systému 2100 Bioanalyzer za využití kitu DNA 1000 (Agilent Technologies) dle doporučení výrobce (Obr. 17).



Obrázek 17: Kapilární elektroforézy směsí produktů mPCR A) před velikostní selekcí B) po velikostní selekci.

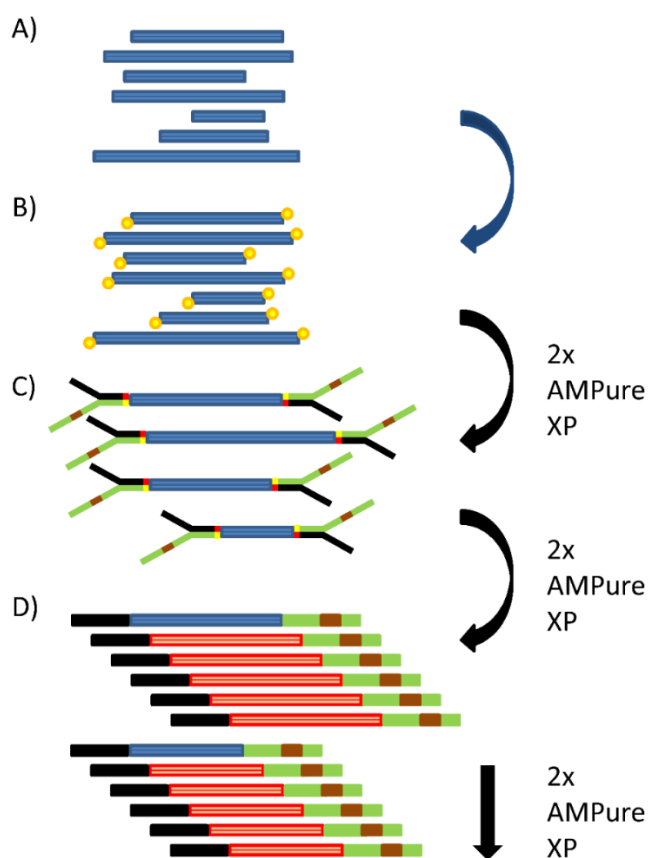
Červenou přerušovanou čarou je vyznačena oblast s amplikony vzniklými na základě alternativně sestřiženého templátu (50–150 bp).

Modrou přerušovanou čarou je vyznačena oblast ampliconů vzniklých pravděpodobně na základě kanonického templátu, které byly odstraněny pomocí velikostní selekce.

Jednotlivé barevné křivky na elektroforegramech odpovídají skupinám vzorků (viz Obr. 15). Vysoké signály v místě 15 a 1500 bp jsou zrovňovací markery.

3.4.3 Příprava sekvenační knihovny

Sekvenační knihovna pro NGS byla připravena z každé ze 12 skupin přečištěných a obohacených vzorků (Obr. 18) dle schématu:



Obrázek 18: Schéma přípravy sekvenační knihovny.

A) Produkty mPCR reakce obohacené o amplikony v rozmezí 50–150 bp.

B) Vytvoření 3'-dA přesahů pomocí Dynazyme II DNA polymerázy.

C) T-A ligace sekvenačních adaptorů (černo-zelený Y komplex) s indexovými sekvencemi pomocí ligačního kitu Rapid DNA.

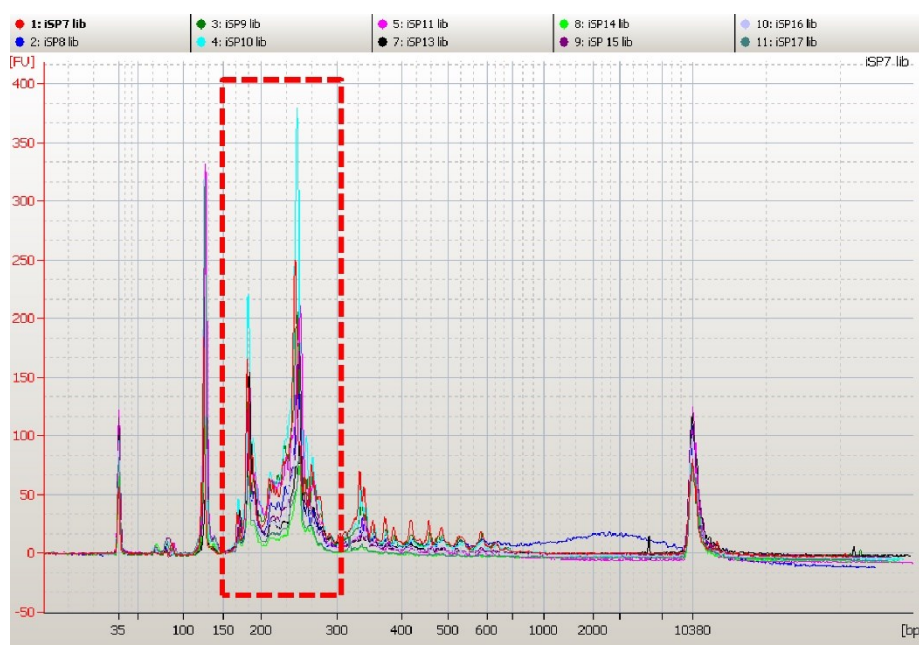
D) 7 cyklů PCR pro namnožení materiálu s pomocí NEBNext High Fidelity PCR Master a primerů s unikátní indexovou sekvencí, které jsou komplementární ke koncovým adaptorovým sekvencím.

Černé šipky značí krok purifikace vzorku pomocí Agencourt AMPure XP systému.

Pro vytvoření sekvenačních knihoven byla nejprve použita Dynazyme II DNA polymeráza (Thermo Fisher Scientific) k obnovení 3'-dA přesahů na PCR templátech (5 μ l 10x Dynazyme buffer; 0,5 μ l Dynazyme II DNA polymerase; 1 μ l 10mM dATP a 43,5 μ l vzorku po dobu 15 min při 72 °C). Dále byly k těmto templátům naligovány Illumina sekvenační adaptory, obsahující indexové sekvence unikátní pro každý vzorek, za využití ligačního kitu Rapid DNA (Thermo Fisher Scientific) dle doporučení výrobce (12 μ l 5x Rapid Ligation Buffer; 1,5 μ l T4 DNA Ligase 5U/ μ l; 1,6 μ l 15 μ M směsi adaptorů; 4,9 μ l ddH₂O a 40 μ l vzorku s A-přesahy po dobu 30 min při 20 °C). Na závěr proběhlo sedm cyklů PCR (98 °C - 30s; 7x 98 °C - 10s → 65 °C - 30s → 72 °C - 30s; 72 °C - 5 min) za využití NEBNext High Fidelity PCR (New England Biolabs) a univerzálními sekvenačními primery v objemu 50 μ l

(22,5 µl vytvořené knihovny; 25 µl NEBNext High Fidelity 2x PCR Master Mix a 2,5 µl 5x PCR primerů).

Po každé enzymatické reakci v přípravě sekvenační knihovny byl vzorek přečištěn pomocí Agencourt AMPure XP Reagent (Beckmann Coulter) s koncentrací směsi 2x. Kvalita připravených knihoven byla charakterizována pomocí kapilární gelové elektroforézy na systému 2100 Bioanalyzer a využití kitu HS DNA (Agilent Technologies) dle doporučení výrobce (Obr. 19). Kvantifikace knihoven proběhla fluorimetricky (Qubit, Thermo Fisher Scientific) dle návodu výrobce.



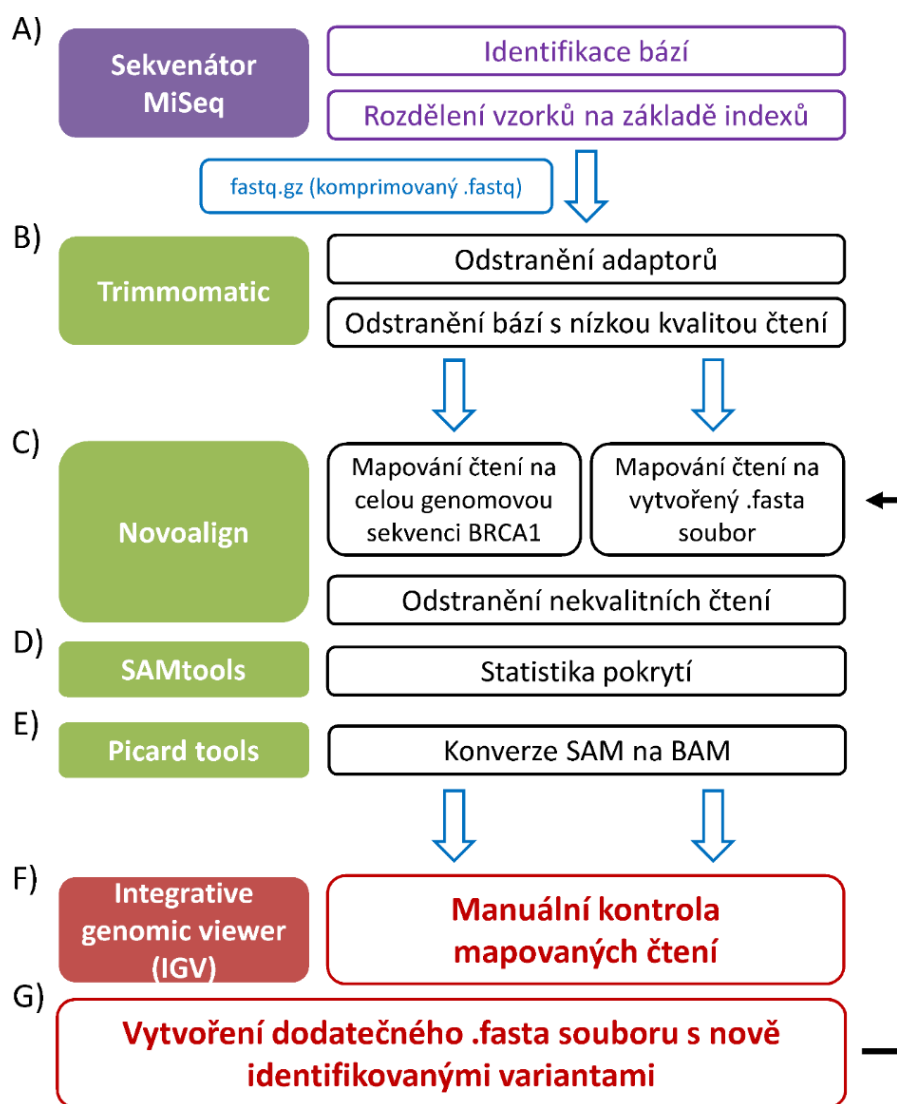
Obrázek 19: Kontrolní kapilární elektroforéza připravených sekvenačních knihoven. Na rozdíl od velikostní selekce byla kontrola provedena pomocí HS (High Sensitivity) DNA kitu, který měří DNA fragmenty v rozmezí cca 50 – 10 000 bp a vysokou citlivostí. Oblast s amplikony našeho zájmu je vyznačena červenou přerušovanou čarou.

3.4.4 Sekvenování na platformě MiSeq

Sekvenace připravených knihoven byla provedena na přístroji MiSeq (Illumina) za použití sekvenačního kitu MiSeq Reagent Kit v3 (150 cyklů) v módu pair-end čtení (75 bází z každé strany DNA inzertu). Pro zajištění dostatečné komplexity sekvenačního běhu byly připravené knihovny kombinovány s odlišnou, panelovou knihovnou tak, aby jedna knihovna našeho zájmu, vytvořená z mPCR směsí, zabírala přibližně 1/30 sekvenační kapacity přístroje.

3.4.5 Biostatistické zpracování dat

Primární hrubá sekvenační data (ve formátu fastq.gz) byla procesována rutinní bioinformatickým postupem vyvinutým v naší laboratoři dle schématu na Obr. 20.



Obrázek 20: Schéma analýzy sekvenačních dat. **A)** Primární automatické zpracování v rámci sekvenačního běhu v přístroji MiSeq. **B)** Odstranění adaptorů a bází s nízkou kvalitou čtení za využití software Trimmomatic. **C)** Paralelní mapování a odstranění nekvalitních čtení pomocí software Novoalign. **D)** Vytvoření základní statistiky pokrytí pomocí software SAMtools. **E)** Převod výstupního formátu SAM ze software Novoalign na formát BAM pomocí software Picard tools. **F)** Manuální kontrola mapovaných čtení v prohlížeči čtení IGV, úprava statistického výstupu (krok D) dle zkontrolovaných dat. **G)** Vytvoření nového vzorového .fasta souboru na základě nově identifikovaných variant (z mapování čtení na celou sekvenci genu), reanalýza těchto variant z důvodu unifikace výsledků. **Fialově** – kroky automatického procesu v rámci sekvenačního běhu MiSeq. **Zeleně** – kroky v laboratoři vytvořeného postupu pro zpracování dat. **Hnědě** – kroky manuální kontroly.

Z hrubých dat byly nejprve odstraněny adaptorové sekvence a nízkokvalitní identifikované báze za využití software Trimmomatic (ver. 0.32 <http://www.usadellab.org>).

Nejprve byla sekvenační data mapována na připravený .fasta soubor pomocí Novoalign (ver. 2.08.03; Novocraft). Souběžně s mapováním byla odstraněna čtení s nedostatečnou sekvenační kvalitou. Připravený „BRCA1 sestřihový“ .fasta soubor obsahoval sekvence, se všemi teoreticky možnými exon-exonovými spojeními všech teoreticky možných sestřihových variant transkriptu BRCA1 (NM_007294) zahrnující alternativní exon 13A a známé varianty exonových variací (viz kapitola 3.4.1.1 *Návrh multiplexní PCR*). Výsledek mapování ve formátu SAM byl konvertován do souboru BAM pomocí Picard Tools (ver. 1.129; <https://broadinstitute.github.io/picard/>). BAM soubory byly kontrolovány v prohlížeči IGV (Integrative Genomics Viewer, Broad Institute) a statistika pokrytí byla vytvořena pomocí SAMtools (ver. 0.1.19; <http://samtools.sourceforge.net/>).

Tento přístup ignoruje přítomnost exonizovaných intronových sekvencí, proto byla hrubá sekvenační data znovu mapována na celý gen *BRCA1* (81,189 bp z referenční NG_005905 mezi 92,500–173,688 bázi). Výsledky mapování byly analyzovány manuálně v prohlížeči IGV. U čtení, která zasahovala z exonu do přilehlého intronu nebo u čtení, která se mapovala uvnitř intronu, byla zaznamenána celá sekvence příslušného čtení, včetně nenamapovaných (skrytých) nukleotidů. Nenamapované části těchto čtení byly porovnávány pomocí BLAST (<https://blast.ncbi.nlm.nih.gov/>). Pokud nenamapované části sekvencí obsahovaly část jiného exonu *BRCA1*, bylo možno identifikovat a popsat celou intronovou inzerci. Celé identifikované sekvence s intronovými částmi byly sestaveny do nového .fasta souboru a použity pro nové mapování za použití originálních hrubých dat stejným principem jako původně.

Všechny výstupy mapovaných čtení byly manuálně zkontrolovány v IGV z důvodu vyloučení chybných mapování. Pro semi-kvantitativní porovnání identifikovaných exon-exonových spojení napříč skupinami byly počty čtení jednotlivých variant vztaženy k 10^6 čtení daného vzorku (směsi vzorků).

Přítomnost volných sestřihových míst byla identifikována pomocí IGV, kde se tato místa zobrazovala jako inserce či delece. Jejich počty byly manuálně odečteny a poměrově zhodnoceny (viz kapitola 4.2.7. *Kvantitativní posouzení variant s volnými sestřihovými místy*).

3.5 Kvantifikace exprese vybraných nalezených ASV BRCA1 pomocí RT-qPCR

Kvantitativní RT-PCR v reálném čase (RT-qPCR) umožňuje velmi přesnou kvantifikaci genové exprese na úrovni mRNA (resp. cDNA). Metoda je založena na PCR reakci využívající měřitelného nárůstu fluorescence emitované fluorochromy přítomnými v reakční směsi, které jsou schopné nespecifické vazby do dvoušroubovice DNA. Hladina fluorescence je po každém cyklu měřena a narůstá přímo úměrně s přibývajícím množstvím produktu, efektivitou PCR reakce a množstvím templátů vstupní mRNA. Výslednou hodnotou qPCR reakce je tzv. „*Crossing point*“ (Cp, někdy také označováno Cq nebo Ct), tato hodnota odpovídá cyklu PCR reakce, ve kterém amplifikační křivka amplikonu protнула hodnotu základní linie (linie šumu). Zpětně je pak vypočítán vstupní poměr sledovaného amplikonu oproti amplikonu referenčnímu za využití znalosti efektivit PCR reakcí obou sledovaných amplikonů, nebo absolutní množství templátů ve vzorku.

3.5.1 Návrh primerů pro RT-qPCR analýzu

Primery pro RT-qPCR analýzu byly specificky navrženy tak, aby byly amplifikovány pouze unikátní sestřihové varianty. Tohoto bylo docíleno navržením jednoho primeru z primerového páru do unikátního exon-exonového spojení. „Přechodový“ primer musel vykazovat vysokou specifitu ke konkrétnímu unikátnímu exonovému spojení. Např. FWD „přechodový“ primer BRCA1Δ9_10 pro transkript s výpadkem exonů 9 a 10 nasedal svou delší částí (5' konec) na konec exonu 8 a kratší 3' částí (3-6 bází) na exon 11. Krátká část „přechodového“ primeru nesměla mít žádnou výraznou shodu se sekvencí exonu ve wt mRNA transkriptu (na tomto příkladu exonu 9). Druhý primer z primerového páru pak nasedal ve stejném exonu v optimální vzdálenosti (Obr. 21).



Obrázek 21: Schéma návrhu primerů pro transkripčně specifickou RT-qPCR. Modře je znázorněn exon A, červeně exon C. Kanonický exon B není v transkriptu přítomen, byl vystřižen pomocí alternativního sestřihu. Černě je znázorněn primerový pár specifický pro alternativní sestřihovou variantu ΔB, kdy levý primer nasedá do unikátního exon-exonového spojení.

Do unikátních exonových spojení bylo navrženo 7 primerových párů, 2 primerové páry byly navrženy pro zhodnocení celkové exprese BRCA1 (na 5' konci v exonu 7 a na 3' konci do exonů 23 až 24). Jeden primerový pár byl navržen pro odlišení krátké formy exonu 11, další byl navržen specifický pro alternativní produkt genu BRCA1 – BRCA1 IRIS. Do každého z referenčních genů (*UBC* – ubikvitin C, NM_021009.6; *ATP5B* – beta podjednotka F1 ATP syntázy NM_001686.3 a *GAPDH* – glyceraldehyd-3-fosfát dehydrogenáza, NM_002046.6) byl navržen jeden primerový pár. Jeden navržený primerový pár byl používán pouze specificky při preamplifikaci oblasti BRCA1 exonu 2 až 6 (na 5' konci transkriptu BRCA1 byly koncentrovány tři varianty $\Delta 3$, $\Delta 3_5$ a $\Delta 5$). Pro kontrolu nulové kontaminace vzorku DNA byl využit jeden primerový pár nasedající do intronových oblastí *CHEK2* (viz příloha 2).

Všechny primery byly navrženy na anelační teplotu v rozmezí (54,3 – 59,8 °C), byla provedena predikce tvorby vlásenek, primerových dimerů a dále byla provedena optimalizace anelačních teplot (T_{an}) - za optimální byla zvolena 62 °C. Dále byla provedena kontrola specifčnosti primerů pomocí 2% gelové elektroforézy a Sangerova sekvenování. U „sestřihových“ primerů byla kontrola provedena amplifikací daných fragmentů na plazmidu obsahující sekvenci kanonického transkriptu BRCA1 mRNA. Pokud se tvořil PCR produkt se „sestřihovými“ primery na kanonickém templátu BRCA1, primery byly znovu navrženy tak, aby vykazovaly vyšší specifitu pro unikátní exon-exonové spojení (viz kapitola 4.3.2. *Optimalizace qPCR reakce*).

Na základě dřívějších zjištění (Lu et al., 1996; viz kapitola 1.3.3. *Alternativní sestřihové varianty BRCA1 a jejich význam*) byly navrženy primery pro variantu $\Delta 9_10$ tak, aby bylo možné tuto variantu kvantifikovat v kombinaci jak s krátkou verzí exonu 11 (11q), tak s dlouhou, kanonickou verzí stejného exonu.

3.5.2 Preamplifikace

Před vlastní qPCR analýzou byla, z důvodu omezeného množství genetického materiálu a relativně nízké exprese mRNA BRCA1, provedena preamplifikace všech vyšetřovaných ASV variant u všech vzorků. Díky preamplifikaci byla zvýšena citlivost detekce a bylo tak možno hodnotit i nízko exprimované varianty. Bez preamplifikace se „crossing“ point (C_p) ampliconu odrážejícího celkovou expresi *BRCA1* přibližoval k číslu C_p (cyklu) 33–35, tedy k detekčnímu limitu metody. Bez preamplifikace by nebylo možné detekovat a analyzovat varianty s nízkou expresí (< 10 % celkové exprese *BRCA1*).

Každý vzorek byl preamplifikován s mixem primerů pro každý amplikon (finální koncentrace v reakci byla 112,5 nM pro každý primer) a pomocí 5x HOT FIREPol Probe qPCR Mix Plus (Solis BioDyne) dle pokynů výrobce ve 20 µl reakci a protokolu:

95 °C	15 min	}	14x
95 °C	15 s		
62 °C	4 min		

Preamplifikované vzorky byly 21x naředěny ddH₂O.

3.5.3 RT-qPCR analýza a analýza křivek tání

Pro vlastní qPCR analýzu byl využit enzym s pufrém obsahujícím interkalační fluorescenční barvivo Eva Green (5x HOT FIREPol EvaGreen qPCR Supermix; Solis BioDyne). Analýza qPCR byla provedena dle doporučení výrobce v reakci o objemu 10 µl, která obsahovala 2 µl 5x amplifikačního mixu, 1 µl mixu obou příslušných primerů (finální koncentrace 175 nM každý), 1 µl preamplifikovaného templátu a 6 µl ddH₂O. Reakce byly pipetovány v 384 jamkovém formátu s využitím pipetovací stanice Pipetmax (Gilson) a analyzovány na stroji LightCycler 480 II (Roche) za využití protokolu:

95 °C	12 min	}	40x
95 °C	10 s		
62 °C	20 s		
72 °C	10 s		

Všechna měření byla prováděna v technických replikátech a specifčnost všech analyzovaných amplikonů byla posouzena v programu LightCycler 480 software (Roche) pomocí analýzy křivky tání (kontinuální zvyšování teploty z 68 °C do 95 °C za neustálého měření klesající fluorescence), která proběhla ihned po vlastní RT-qPCR analýze. V případě detekované odchylky křivky tání byla provedena 2% gelová elektroforéza produktu a Sangerovým sekvenováním zjištěna sekvence qPCR produktu.

Standardně byly prováděny negativní kontroly qPCR (obsahující stejný mix, pouze templátová cDNA byla nahrazena ddH₂O).

3.5.4 Stanovení efektivity qPCR reakcí

Z ředících řad cDNA byly zkonstruovány za optimalizovaných reakčních podmínek kalibrační křivky pro všechny vyšetřované amplikony. Z nich byla poté stanovena efektivita jednotlivých RT-qPCR reakcí. Pro konstrukci kalibračních křivek bylo z buněčné linie MCF7 použito 200 ng, 100 ng, 50 ng, 25 ng, 10 ng, 1 ng a 0,1 ng izolované RNA (cDNA resp.) na 10 µl reakci (Hojný J., 2012).

Kalibrační křivky pro absolutní kvantifikaci byly sestaveny za využití specifických, arteficiálních PCR produktů všech charakterizovaných amplikonů (včetně unikátních sestřihových variant). Spektrofotometricky (NanoDrop 1000) byla určena jejich koncentrace a po 10x naředění a byla v tripletu stanovena finální koncentrace, která byla použita do přepočtu počtu templátů na µl podle následujícího vzorce:

$$\text{Konc. vzorku (molekuly}/\mu\text{l}) = \frac{(\text{koncentrace vzorku [ng}/\mu\text{l}] \times 6,022 \times 10^{23})}{656,6 \times 10^9 \times a}$$

kde *a* označuje délku PCR produktu v bp.

Po výpočtu byly všechny arteficiální PCR produkty naředěny na koncentraci 10¹⁰ templátů/µl. Byl připraven směsný roztok všech cílových sekvencí, ve kterém měla každá sekvence koncentraci 10⁹ templátů/µl. Výsledný pracovní roztok qPCR templátů sloužil pro diluční kalibrační řadu o koncentracích 1x10⁵, 1x10⁴, 1x10³, 5x10², 2,5x10², 1x10², 50, 25, 10, 1 každého z templátů/µl, která byla použita pro konstrukci kalibračních křivek qPCR a zjištění detekčního limitu metody (Hojný J., 2012).

3.5.5 Kvantifikace změn exprese jednotlivých ASV a jejich statistické vyhodnocení

Výstupem qPCR analýzy byly hodnoty Cp jednotlivých amplikonů, které nezohledňují efektivitu jednotlivých reakcí. Hrubá data byla nejprve zpracována pomocí softwaru qGENE (<http://www.qgene.org>), který je aplikací do tabulkového kalkulátoru Excel (Microsoft) a pomocí software GenEX (MultiD). Oba programy umožňují efektivitu jednotlivých qPCR reakcí zohlednit.

Vstupní množství templátové cDNA bylo normalizováno a míra relativní exprese BRCA1 ve vyšetřovaných tkáních byla posuzována v porovnání s referenčními geny pro ubikvitin C (*UBC*), katalytickou podjednotku beta F1 ATP syntázy (*ATP5B*) a glycerinaldehyd 3-fosfát

dehydrogenázu (*GAPDH*), které byly vybrány na základě předchozích pokusů prováděných na pracovišti pro tento typ analýz a na základě dokumentované exprese těchto referenčních genů v prsní tkáni (viz kapitola 4.3.3. *Zhodnocení vybraných referenčních genů*). Hodnocení míry exprese sledovaných ASV BRCA1 bylo normalizováno na míru exprese BRCA1 exonů 7 a 23/24 (viz kapitola 4.3.4. *Celková exprese BRCA1 ve vyšetřovaných tkáních*; Favy et al., 2000).

Pro hodnocení relativní míry exprese sledovaných amplikonů v jednotlivých typech tkání napříč vyšetřovanými skupinami byl použit software GenEX v6.1 (MultiD), stejně jako pro tvorbu popisných statistik jednotlivých souborů. Software GenEX je specializovaným RT-qPCR nástrojem, který kromě základního zpracování a vyhodnocení dat umožňuje i tvorbu pokročilých statistických analýz na úrovni souborů. Software GenEX byl vytvořen v souladu s MIQE (Bustin et al., 2009) a CLSI (Clinical Laboratory Standards Institute; <https://clsi.org/>) doporučeními.

Změny relativní exprese u sledovaných sestřihových variant v souborech pacientů s karcinomem prsu a zdravých kontrol a jejich dynamika, byla vyhodnocena pomocí statistického software REST 2009 ver. 2.0.13 (Relative Expression Software Tool; <https://www.gene-quantification.de/rest-2009.html>; Corbett Research Pty. Ltd; Pfaffl, 2001; Pfaffl et al., 2002) a pomocí software GenEX.

Pro statistické zpracování souborů byl využit neparametrický Mann-Whitney U-test (v rámci software GenEX). Z důvodu omezení statistické chyby prvního typu v rámci Mann-Whitney U-testu byly za signifikantní výsledky považovány výsledky s $p < 0,0085$ (6 porovnávaných souborů) a $p < 0,0073$ (7 porovnávaných souborů).

Pro tvorbu grafů byl využit program Excel (Microsoft). Krabicové grafy jsou standardně sestaveny na základě hodnot 1. a 3. kvartilu s vnitřní hodnotou druhého kvartilu (medián), chybové úsečky značí minimum a maximum, odlehle hodnoty jsou znázorněny tečkou a průměr je znázorněn křížkem.

Korelační analýzy pro porovnání exprese jednotlivých variant mezi vyšetřovanými tkáněmi a porovnání exprese v pravé a levé prsní tkáni včetně vyhodnocení statistické signifikance byly provedeny pomocí software Statistica 12 (StatSoft) pomocí funkce korelační matice, stejně jako lineární regrese varianty $\Delta 9_{10}$ u skupin pacientek s karcinomem prsu a zdravých kontrol v závislosti na věku, která byla vypočítána pomocí funkce lineární rovnice.

4 VÝSLEDKY

4.1 Izolace a zpracování biologického materiálu

4.1.1 Izolace a kontrola kvality celkové RNA z lidských tkání

Řada vzorků od celkově 216 žen byla v průběhu zpracování a následných analýz vyloučena. I přes použití RNA prezervačních médií a přísných metodických postupů se, především z důvodu nedostatečně kvalitního tkáňového materiálu (nízká buněčnost) a lidské chyby (vysoká míra degradace RNA, chyby při izolaci RNA apod.), nepodařilo v některých případech izolovat dostatečné množství kvalitní RNA.

Úspěšně a s dostatečným výtěžkem se pro syntézu cDNA podařilo izolovat 633 vzorků z celkových 854 vzorků (74,1 %). Z těchto vzorků mělo dostatečnou integritu RNA (RIN > 5) 596 (94,2 %). Všechny vzorky s dostatečnou integritou RNA úspěšně prošly syntézou cDNA.

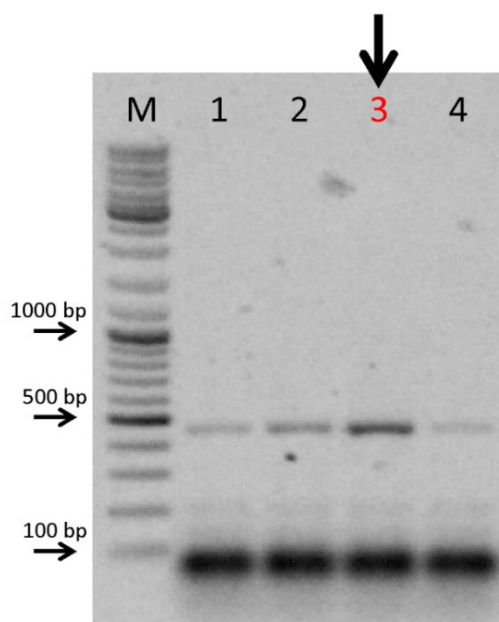
4.1.2 Optimalizace reverzní transkripce

Pro syntézu cDNA byla použita reverzní transkriptáza SSIII (Thermo Fisher Scientific) s náhodnými hexanukleotidy. V průběhu reverzní transkripce s hexanukleotidy vzniká po celé délce transkriptu množství cDNA úseků, jejichž délka závisí na koncentraci hexanukleotidů v reakci, čím menší je koncentrace hexanukleotidů, tím delší úseky se syntetizují. Výsledná cDNA je za použití hexanukleotidů kvantitativním obrazem původní RNA, protože jsou pokryty všechny úseky transkriptu. V průběhu optimalizace bylo porovnáváno množství a dostatečná délka PCR produktů vzniklých z přepsané cDNA (Tab. 5).

Tabulka 5: Optimalizace reverzní transkripce. Červeně je vyznačeno složení směsi, které bylo kontrolní PCR vyhodnoceno jako optimální pro nejlepší výsledek reverzní transkripce. Hodnoty v tabulce jsou uvedené v μl .

	1	2	3	4
RNA ošetřena DNázou (1 μg)	6,1	6,1	6,1	6,1
Hexanukleotidy (30 pmol/ μl)	0,25	0,5	0,25	0,5
ddH ₂ O	0,25	-	0,5	0,25
50x dNTPs	0,4	0,4	0,4	0,4
DTT	0,5	0,5	0,5	0,5
5x buffer	2	2	2	2
SS3	0,5	0,5	0,25	0,25

Kontrolní PCR reakce proběhly pomocí FastStart Taq DNA Polymerase kitu (Roche) dle doporučeného protokolu výrobce v objemu 10 μ l a jako templát byl použit 1 μ g celkové RNA (z buněčné linie MCF-7), která byla těsně před reverzní transkripcí ošetřena DNázou I (viz kapitola 3.3.6. *Syntéza cDNA*). Jako vhodné primery pro kontrolní PCR byly vybrány primery HTS e5F a HTS e10R, které dávají vzniknout ampliconu o dostatečné velikosti 467 bp (Obr. 22).



Obrázek 22: Optimalizace reverzní transkripce. Jednotlivé kombinace množství reverzní transkriptázy a hexanukleotidů z tabulky V1 jsou vyznačeny čísly 1–4. M – velikostní standard (marker). Černou vertikální šipkou je označen vzorek s největším výtěžkem amplifikace. PCR produkty jsou zobrazeny na 1% agarózovém gelu.

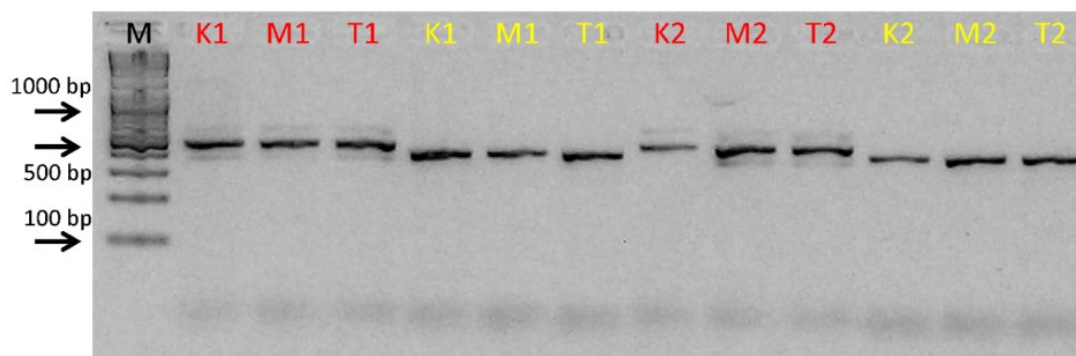
Výsledky kontrolní elektroforézy ukázaly nejvyšší výtěžek kontrolní PCR reakce na templátu cDNA, která byla syntetizována za použití 0,25 μ l náhodných hexanukleotidů a 0,25 μ l SSIII reverzní transkriptázy v reakčním objemu 10 μ l (Obr. 22). V průběhu zpracování lidské RNA se syntéza cDNA prováděla ve dvojnásobném objemu.

4.2 Identifikace alternativních sestřihových variant *BRCA1* v biologických vzorcích

Cílem první části práce bylo zavedení robustního systému pro rychlou a citlivou analýzu alternativních sestřihových variant jakéhokoliv genu, na modelu *BRCA1*, za pomoci multiplexní PCR, obohacení vzorků o krátké amplikony s unikátními exon-exonovými spojeními, přípravy sekvenačních knihoven ze selektovaných vzorků, NGS s vysokým pokrytím a biostatistické analýzy sekvenačních dat.

4.2.1 Kontrola správného průběhu syntézy cDNA pro identifikaci ASV

Pro metodu identifikace ASV *BRCA1* byly vybrány vzorky, které měly nejvyšší možnou kvalitu izolované RNA ve všech třech vyšetřovaných tkáních (průměr RIN = 7,4) a zároveň dostatečné množství izolované RNA (pro zachování genetického materiálu pro následné RT-qPCR analýzy). Kvalita syntetizované cDNA z těchto vzorků byla ověřena kontrolní PCR (viz kapitola 3.3.7 *PCR kontrola reverzní transkripce syntetizované cDNA pro identifikaci ASV*, Obr. 23)

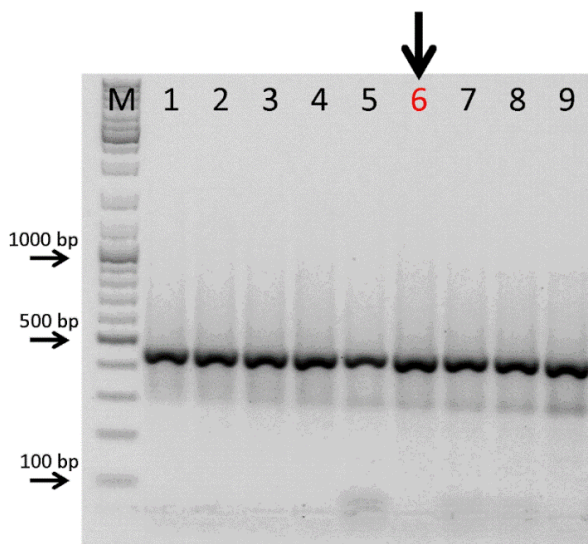


Obrázek 23: Příklad elektroforézy kontrolní PCR syntézy cDNA. **1** – vzorek č. 173. **2** – vzorek č. 186. **K** – RNA izolovaná z leukocyte periferní krve. **M** – RNA izolovaná z mamární tkáně. **T** – RNA izolovaná z přilehlé tukové tkáně. **Červeně** – amplikon s primery v exonech 2 a 8. **Žlutě** – amplikon s primery v exonech 15 a 17. **M** – velikostní marker. PCR produkty jsou zobrazeny na 1,5% agarózovém gelu.

Z celkem 96 vzorků syntetizované cDNA (4 skupiny žen x 3 vyšetřované tkáně x 8 vzorků ve skupině) odhalila kontrolní elektroforéza 3 vzorky bez výsledného PCR produktu. Syntéza cDNA z těchto vzorků byla následně úspěšně zopakována, což potvrdila dodatečná kontrola.

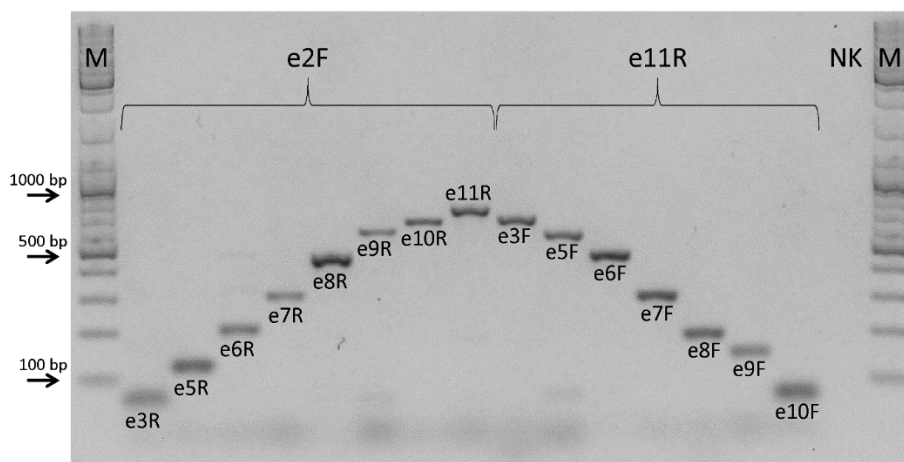
4.2.2 Optimalizace primerů pro multiplexní PCR reakce

U primerů pro amplifikaci exon-exonových spojení byla optimalizována anelační teplota (T_{an}) pro nasedání primerů a ověřena jejich funkčnost a dostatečná specifita. Optimalizace T_{an} probíhaly v objemu 10 μ l (viz kapitola 3.4.1.2 *Vlastní mPCR reakce*), v cykleru PTC-DyAD (BioRad) s gradientovým blokem v rozmezí teplot 56–64 °C. Za optimální T_{an} byla zvolena teplota 62 °C, s přihlédnutím k teplotě vypočítanou *in-silico* softwarem Annhyb (Obr. 24).



Obrázek 24: Optimalizace multiplexní PCR s primery HTS e6F a HTS e11R (440 bp). Reakce s různými anelačními teplotami jsou vyznačeny čísly 1–9 (1 – 56 °C; 2 – 57,1 °C; 3 – 58,3 °C; 4 – 59,6 °C; 5 – 60,8 °C; 6 – 61,9 °C; 7 – 62,7 °C; 8 – 63,2 °C; 9 – 64 °C). M – velikostní standard (marker). Černou vertikální šipkou je označen vzorek s nejvyšší účinností amplifikace při anelační teplotě primerů 62 °C. PCR produkty jsou zobrazeny na 1% agarózovém gelu.

Po optimalizování anelační teploty byl proveden test amplifikovatelnosti se všemi navrženými primery, a všemi jejich možnými kombinacemi pomocí kontrolních PCR reakcí (viz příloha 1; Obr. 25.).

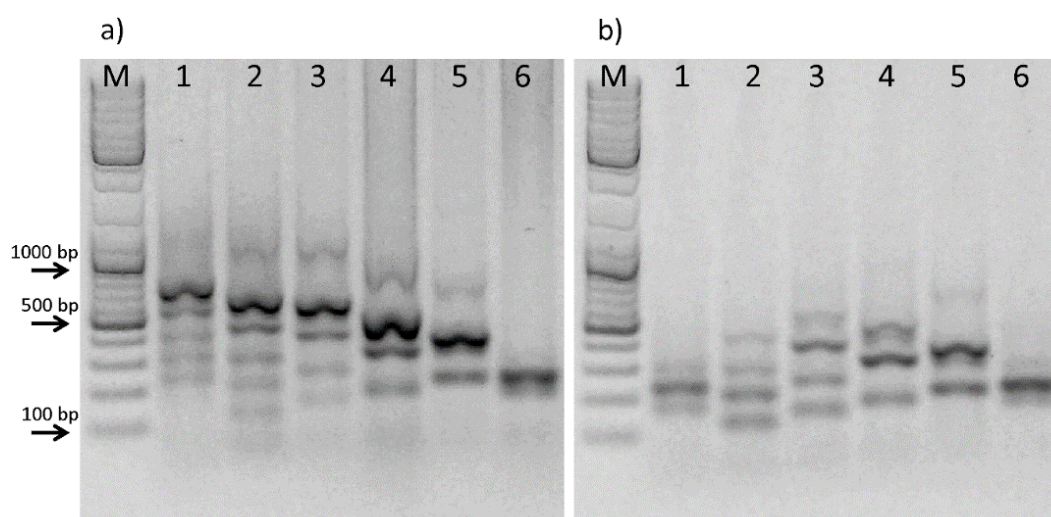


Obrázek 25: Test funkčnosti HTS primerů na příkladu primerů e2F a e11R a jejich partnerů (viz kapitola 3.4.1.1 *Návrh multiplexní PCR (mPCR)*, Obr. 14). M – velikostní standard (marker). NK – negativní kontrola. PCR produkty jsou zobrazeny na 1% agarózovém gelu.

Pro testování byl použit plazmid pcDNA3.1 s inkorporovanou kódující oblastí transkripční varianty BRCA1 11q. Tento konstrukt sloužil jako templát pro kontrolní PCR reakce z důvodu usnadnění amplifikace a snazší interpretace výsledků – jeden primerový pár dá vzniknout pouze jednomu produktu (Obr. 25; BRCA1 11q je variantou BRCA1 se zkrácenou formou exonu 11 o 3309 bp; tento konstrukt byl v naší laboratoři k dispozici z předchozích experimentů). Detekované nefunkční primery (e11F, e12R, e15R) byly upraveny a znovu testovány. Primery pro známé alternativní varianty (zkrácená forma exonu 5 – 5q, alternativní exon 13 – 13A), primer nenasedající do kódující oblasti (e1F) a primer nasedající do oblasti dlouhé formy exonu 11 (která nebyla v použitém plazmidu přítomna; e11F) byly otestovány na cDNA templátu syntetizované z celkové RNA buněčné linie MCF-7, ve které byly všechny tyto varianty přítomny.

4.2.3 Optimalizace průběhu multiplexních PCR reakcí

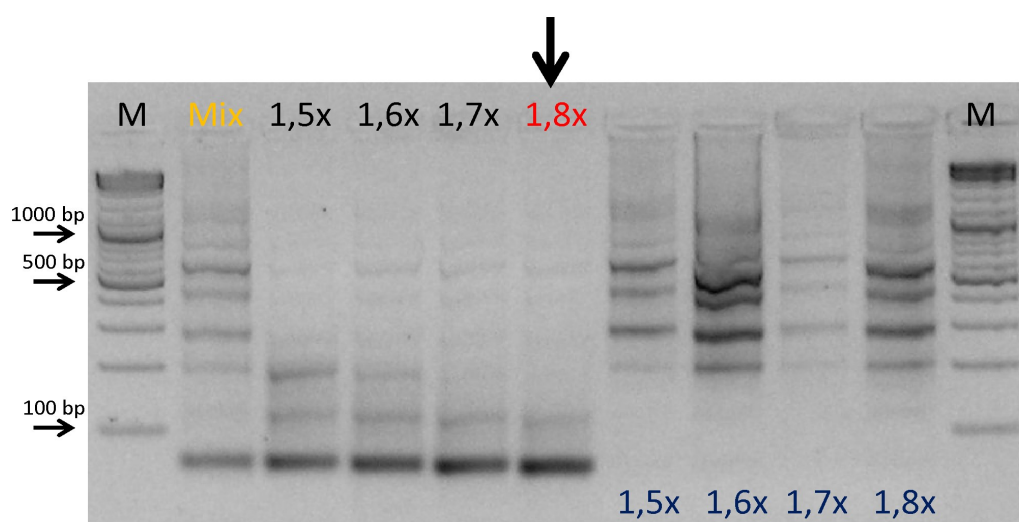
Délka doby polymerace (elongace polymerázy) v průběhu PCR reakce byla optimalizována, aby byla omezena amplifikace dlouhých templátů vzniklých z kanonického BRCA1 templátu. Čas elongace PCR polymerázy byl v průběhu optimalizace snižován (Obr. 26). Jako výsledný čas elongace byl vybrán hraniční čas 15 s.



Obrázek 26: Optimalizace času elongace během mPCR. a) Prvních 6 mPCR reakcí s délkou elongace 1 minuta. b) Prvních 6 mPCR reakcí s délkou elongace 15 vteřin. Čísly 1–6 jsou označeny jednotlivé mPCR reakce (viz kapitola 3.4.1.1 *Návrh multiplexní PCR (mPCR)*). M – velikostní standard (marker). PCR produkty jsou zobrazeny na 1% agarózovém gelu.

4.2.4 Optimalizace velikostní selekce produktů mPCR reakcí

Amplikony o velikostech větších než 200 bp, které v produktech mPCR stále přetrvávaly (produkty kanonického BRCA1 templátu), měly být odstraněny pomocí systému AMPure XP (Beckmann Coulter). Na základě různé finální koncentrace polyethylenglykolu (PEG; v roztoku AMPure XP) a vzorku je ovlivněna afinita dlouhých DNA řetězců k porézním magnetickým partikulím (čím vyšší koncentrace PEG, tím kratší DNA řetězce se na magnetické partikule váží; Obr. 27). Pro odstranění ampliconů > 150 bp jsme zvolili 1,8x vstupní koncentrace PEG do reakce.



Obrázek 27: Optimalizace velikostní selekce. Mix – směs jednotlivých mPCR reakcí (neselektovaný vzorek). Jednotlivá čísla udávají vstupní koncentrace PEG, černými nápisy jsou označeny selektované části vzorků našeho zájmu a modře části vzorků, které byly odstraněny. Černou vertikální šipkou je označen vybraný vzorek s optimální účinností velikostní selekce. M – velikostní standard (marker). PCR produkty jsou zobrazeny na 2% agarózovém gelu.

4.2.5 Souhrnné výsledky mapování ASV BRCA1

Mapování veškerých sekvenačních dat (viz kapitola 3.4.5. *Biostatistické zpracování dat*) s následnou kontrolou výsledků odhalila přítomnost 94 různých alternativních sestřihových variant BRCA1 napříč všemi vyšetřovanými vzorky (Tab. 6).

Tabulka 6: Spektrum všech nalezených alternativních sestřihových variant BRCA1. Čísla vyjadřující množství variant jsou normalizována na 10⁶ čtení. Barevným polem je vyznačena četnost nalezené varianty: **Bíle** – 0; **Zeleně** – 1-9; **Žlutě** – 10-99; **Světle červeně** – 100-999; **Červeně** – 1000-9999; **Tmavě červeně** > 9999. **FS** značí „frameshift“; **IF** „in-frame“; **n.c.** – nekódující variantu. **C** = kazetová; **mC** = multikazetová sestřihová událost. **SDS/SAS** – donorové / akceptorové sestřihové místo; **Δ** = delece; **▼** = inserce. Tučně s modrým pozadím jsou označeny varianty vyskytující se ve všech vyšetřovaných lidských tkáních. Sloupec **Reference** značí, zda byla varianta identifikována v minulosti: **0** = ne; **1** = ano; **P** = byla označena jako predominantní.

Popis varianty	HGVS nomenklatura	Funkční anotace	Biotyp	Test. buněčné linie		Analyzované lidské tkáně									Reference					
				MCF7	mix	Leukocyty			Mama			Tuková tk.			Leu		Mam	Tu		
						Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Colombo, 2014	Romero, 2015	Colombo, 2014	Romero, 2015	Romero, 2015	Orban Olah, 2003
1Aq	c.-25_-20del6	UTR	SDSΔ	1185	7854	3632	5045	4210	2981	11439	4010	2508	6035	4955	P	P	P	P	P	1
1Aq, 2a	c.-25_-20del6, c.-19-59_-19-1 ins59	UTR	SDSΔ + SAS▼	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2	c.-25_80del105	n.c.	SDSΔ + CΔ	176	5032	40	101	150	46	191	28	59	82	205	1	1	1	0	0	0
1Aq, Δ2_3	c.-25_134del159	n.c.	SDSΔ + mCΔ	0	1575	0	202	13	14	46	0	0	0	0	1	1	1	0	0	0
1Aq, Δ2_5	c.-25_212del237	n.c.	SDSΔ + mCΔ	0	438	0	0	0	2	0	83	0	0	0	1	1	1	0	0	0
1Aq, Δ2_5, 6p	c.-25_282del307	n.c.	SDSΔ + mCΔ + SASΔ	0	350	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_7	c.-25_441del466	n.c.	SDSΔ + mCΔ	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_7, 8p	c.-25_444del469	n.c.	SDSΔ + mCΔ + SASΔ	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_10	c.-25_670del695	n.c.	SDSΔ + mCΔ	0	438	0	0	7	0	0	0	0	0	0	1	1	1	0	0	0
1Aq, Δ2_17	c.-25_5074del5099	n.c.	SDSΔ + mCΔ	88	1225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_19	c.-25_5193del5218	n.c.	SDSΔ + mCΔ	0	0	0	0	0	0	289	0	0	0	0	0	0	0	0	0	0
1aA	c.-20+1_-20+89ins89	UTR	SDS▼	6824	12098	321	219	1211	1501	6557	5429	940	427	349	1	1	1	0	P	0
2p	c.-19_-7del13	UTR	SASΔ	88	963	140	118	232	77	214	28	76	238	164	1	P	1	0	P	0
Δ2	c.-19_80del99	n.c.	CΔ	176	3982	482	1063	1987	323	3305	1337	481	1308	462	1	P	1	P	P	0
Δ2_3	c.-19_134del153	n.c.	mCΔ	0	1575	0	0	0	36	364	0	5	0	82	1	P	1	0	P	0
Δ2_3, ▼4	c.-19_134del153 + c.135-4047_135-3932ins116	n.c.	mCΔ + C▼	110	0	0	0	0	0	0	0	5	0	51	1	1	0	0	0	0
Δ2_5	c.-19_212del231	n.c.	mCΔ	88	1203	20	0	0	73	35	0	0	0	0	1	1	1	0	0	0
Δ2_7, 8p	c.-19_444del463	n.c.	mCΔ + SASΔ	0	2253	10	169	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ2_10	c.-19_670del689	n.c.	mCΔ	0	88	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Δ2_17	c.-19_5074del5093	n.c.	mCΔ	0	2625	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0
Δ2_19	c.-19_5193del5212	n.c.	mCΔ	0	0	0	101	0	0	976	0	0	22	0	0	0	0	0	0	0
▼145 bp int 2	c.81-3486_81-3342ins145	FS	C▼	636	459	40	34	72	24	87	276	86	100	82	0	0	0	0	0	0
Δ3	c.81_134del54	FS	CΔ	987	21089	3120	10293	8436	4141	9336	9438	2135	5217	6750	1	P	1	0	P	1
Δ3, ▼4	c.81_134del54 + c.135-4047_135-3932ins116	FS	CΔ + C▼	22	44	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0
Δ3_5	c.81_212del132	IF	mCΔ	0	875	60	405	215	109	474	165	108	0	0	0	0	0	0	0	0
▼116bp int 3	c.134+3124_134+3239ins116	FS	C▼	461	1663	20	0	26	61	23	83	38	212	62	0	0	0	0	0	0
▼4	c.135-4047_135-3932ins116	FS	C▼	7460	2188	281	489	238	367	306	276	470	279	574	1	1	1	0	P	0
Δ5	c.135_212del78	IF	CΔ	395	7351	8016	13043	8569	5524	7476	10169	4778	4013	8823	P	1	P	P	P	1
Δ5_6	c.135_301del167	FS	mCΔ	0	0	0	0	0	0	0	634	0	0	0	0	0	0	0	0	0
Δ5_6, 7p	c.135_307del173	FS	mCΔ + SASΔ	0	0	0	0	333	0	127	0	0	0	0	0	0	0	0	0	0
Δ5_7	c.135_441del307	FS	mCΔ	0	481	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ5_7, 8p	c.135_444del310	FS	mCΔ + SASΔ	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ5_9	c.135_593del459	IF	mCΔ	0	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

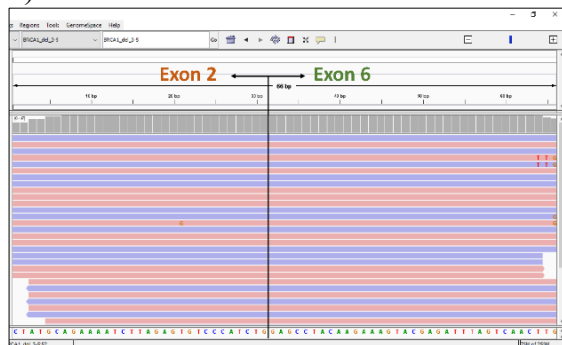
Popis varianty	HGVS nomenklatura	Funkční anotace	Biotyp	Test. buněčné linie		Analýzované lidské tkáně									Reference					
				MCF7	mix	Leukocyty			Mama			Tuková tk.			Leu		Mam		Tu	Orban Olah, 2003
						Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Colombo, 2014	Romero, 2015	Colombo, 2014	Romero, 2015	Romero, 2015	
5q	c.191_212del22	FS	SDSΔ	549	1269	3602	2818	2137	8023	4501	3624	5886	9339	4268	P	P	P	P	P	1
5q, Δ6	c.191_301del111	IF	SDSΔ + CΔ	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	1
Δ6	c.213_301del89	FS	CΔ	44	0	40	0	0	65	35	0	0	19	0	0	0	0	0	0	
6q	c.293_301del9	IF	SDSΔ	0	44	431	371	695	202	116	28	124	7	62	0	0	0	0	0	
8p	c.442_444del3	IF	SASΔ	197	7022	15079	10883	8566	7283	4287	4657	6356	5024	4011	P	P	P	P	P	1
Δ8	c.442_547del106	FS	CΔ	0	0	10	0	0	0	0	55	92	0	0	0	0	0	0	0	0
Δ8_9	c.442_593del152	FS	mCΔ	0	0	0	0	20	141	0	0	0	0	0	1	1	0	0	P	0
Δ8_10	c.442_670del229	FS	mCΔ	0	88	0	0	0	2	0	0	0	0	0	1	1	1	0	P	0
Δ8_16	c.442_4986del4545	IF	mCΔ	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
▼94 bp int 8	c.548-297_548-204ins94	FS	C▼	110	481	251	202	457	71	17	69	238	0	154	0	0	0	0	0	0
▼97 bp int 8	c.548-300_548-204ins97	FS	C▼	461	438	371	456	757	244	312	248	286	204	92	0	0	0	0	0	0
Δ9	c.548_593del46	FS	CΔ	1295	9888	3642	12334	17110	2295	7303	9301	2092	7042	8689	P	1	P	0	P	1
Δ9_10*	c.548_670del123	IF	mCΔ	4542	19558	562	1063	2261	998	2155	3362	940	1561	5109	P	P	P	P	P	1
Δ9_11	c.548_4096del3549	IF	mCΔ	1009	16079	100	270	454	2	0	413	22	0	923	P	1	P	0	0	1
Δ9_12	c.548_4185del3638	FS	mCΔ	0	0	10	0	0	0	0	165	16	0	0	1	1	1	0	0	0
10a	c.594-21_594-1ins21	IF	SAS▼	4827	2056	1575	5113	594	430	4957	1157	665	465	1036	0	0	0	0	0	0
Δ10	c.594_670del77	FS	CΔ	154	1684	30	34	0	12	0	0	27	7	0	1	1	1	0	P	0
Δ10_11	c.594_4096del3503	FS	mCΔ	0	4922	70	101	131	22	0	0	0	0	513	1	1	1	0	0	0
Δ10_12	c.594_4185del3592	FS	mCΔ	0	416	0	0	0	0	0	0	0	89	0	1	1	1	0	0	0
Δ11	c.671_4096del3426	IF	CΔ	439	5163	60	337	104	69	185	303	265	71	246	1	1	1	0	0	1
Δ11_12	c.671_4185del3515	FS	mCΔ	0	0	0	0	0	10	23	0	0	0	0	1	0	1	0	0	0
Δ11_12, 13p	c.671_4188del3518	FS	mCΔ + SASΔ	0	175	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
11q	c.788_4096del3309	IF	SDSΔ	6012	16451	823	928	1965	2971	1491	5883	1048	2557	7879	P	1	P	0	0	1
11 Δ3094	c.788_3881del3094	FS	Intronizatio n	241	219	0	0	39	6	0	0	0	0	0	0	0	0	0	0	0
11 Δ3110	c.788_3897del3110	FS	Intronizatio n	2041	700	40	67	7	32	0	220	32	223	41	1	1	0	0	0	0
11 Δ3240	c.788_4027del3240	IF	Intronizatio n	614	0	80	34	0	0	0	0	0	0	0	1	1	0	0	0	0
13p	c.4186_4188del3	IF	SASΔ	263	1378	3652	3881	9294	2684	560	3114	3297	847	6217	P	1	P	0	P	0
Δ13	c.4186_4357del172	FS	CΔ	44	1006	70	405	111	36	69	138	86	141	482	1	1	0	0	P	0
Δ13, 14p	c.4186_4360del175	FS	CΔ + SASΔ	44	219	20	0	0	6	12	0	0	0	0	1	1	0	0	0	0
Δ13_14	c.4186_4484del299	FS	mCΔ	0	2231	0	675	333	0	0	0	76	0	0	0	0	0	0	0	0
Δ13_15	c.4186_4675del490	FS	mCΔ	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
▼13A	c.4358-2785_4358-2720ins66	IF	C▼	6495	24721	3963	1974	7267	873	1196	703	1254	5953	4309	1	1	1	0	P	0
▼13A, 14p	c.4358-2785_4358-2720ins66 + c.4358_4360del3	IF	C▼ + SASΔ	6604	40494	130	169	166	75	12	14	49	123	21	1	1	1	0	0	0
▼13A, Δ14	c.4358-2785_4358-2720ins66 + c.4358_4484del127	FS	C▼ + CΔ	0	0	0	0	819	115	0	0	0	0	0	0	0	0	0	0	0
14p	c.4358_4360del3	IF	SASΔ	4125	3238	542	1282	917	333	335	358	303	892	533	P	P	P	P	P	0
Δ14	c.4358_4484del127	FS	CΔ	417	1488	1304	877	1429	607	1231	1529	773	1442	882	1	0	0	0	0	0
Δ14_15	c.4358_4675del318	IF	mCΔ	241	1663	10	67	52	44	0	55	38	0	144	1	1	0	0	0	0
Δ14_17	c.4358_5074del717	IF	mCΔ	132	919	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Δ15	c.4485_4675del191	FS	CΔ	285	766	642	911	1723	220	254	468	168	286	1231	1	1	0	0	P	0
Δ15_16	c.4485_4986del502	FS	mCΔ	0	1619	30	0	274	0	0	0	0	0	0	0	0	0	0	0	0
Δ15_17	c.4485_5074del590	FS	mCΔ	176	372	130	945	1214	8	162	55	103	0	718	1	1	1	0	P	1
Δ15_19	c.4485_5193del709	FS	mCΔ	88	744	0	0	228	4	0	0	0	0	0	1	1	1	0	0	0
Δ15_23	c.4485_5467del983	FS	mCΔ	0	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16a	c.4986+1_4986+65ins65	FS	SDS▼	570	44	5568	0	6687	944	2461	3541	740	4292	0	0	0	0	0	0	0
Δ17	c.4987_5074del88	FS	CΔ	88	1466	2358	337	1951	1265	503	7854	1762	1988	6217	1	1	0	0	P	0
Δ17_18	c.4987_5152del166	FS	mCΔ	0	613	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ17_19	c.4987_5193del207	IF	mCΔ	88	1006	0	0	75	0	532	262	0	0	0	0	0	0	0	0	0
Δ17_20	c.4987_5277del291	IF	mCΔ	0	0	0	0	0	107	0	0	0	0	0	0	0	0	0	0	0
Δ18	c.5075_5152del78	IF	CΔ	0	219	60	0	0	44	0	0	0	0	0	1	1	0	0	0	0
Δ18_19	c.5075_5193del119	FS	mCΔ	0	0	10	67	0	4	81	28	11	7	0	0	0	0	0	0	0
Δ18_20	c.5075_5277del203	FS	mCΔ	0	0	30	0	0	4	0	0	0	0	0	0	0	0	0	0	0

Popis varianty	HGVS nomenklatura	Funkční anotace	Biotyp	Test. buněčné linie		Analýzované lidské tkáně									Reference					
						Leukocyty			Mama			Tuková tk.			Leu		Mam		Tu	Orban Olah, 2003
				MCF7	mix	Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Zdravé kontroly	BC pacienti	BRCA1 mut.	Colombo, 2014	Romero, 2015	Colombo, 2014	Romero, 2015	Romero, 2015	
▼143 bp int 19	c.5194-1231_5194-1089ins143	FS	C▼	132	0	20	0	0	381	324	138	573	0	431	0	0	0	0	0	0
▼146 bp int 19	c.5194-1234_5194-1089ins146	FS	C▼	373	197	80	101	653	40	0	289	32	19	421	0	0	0	0	0	0
Δ20	c.5194_5277del84	IF	CA	0	481	191	607	0	60	0	2618	0	0	0	1	1	1	0	0	0
Δ21	c.5278_5332del55	FS	CA	1711	8226	492	776	1765	389	376	1226	195	22	1057	1	1	1	0	P	0
Δ21_22	c.5278_5406del129	IF	mCA	88	438	0	0	0	0	549	0	0	0	0	1	1	1	0	P	0
▼129 bp int 21	c.5332+873_5332+1001ins129	FS	C▼	1711	263	592	641	1250	230	618	400	432	342	657	0	0	0	0	0	0
▼119 bp int 21	c.5333-706_5333-588ins119	FS	C▼	66	328	90	0	65	28	52	179	11	45	318	0	0	0	0	0	0
Δ22	c.5333_5406del174	FS	CA	4739	16451	1585	4590	3064	4405	11919	7372	3464	3872	5160	1	P	1	0	P	0
23a	c.5407-9_5407-1ins9	IF	SAS▼	44	66	10	0	39	35	168	55	68	74	0	0	0	0	0	0	0
Δ23	c.5407_5467del61	FS	CA	263	613	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
Suma identifikovaných variant v každém mPCR mixu vzorků				54	71	56	46	51	64	49	48	50	40	41						
Počet identifikovaných variant v analyzovaných tk. typech				76		67			72			54								

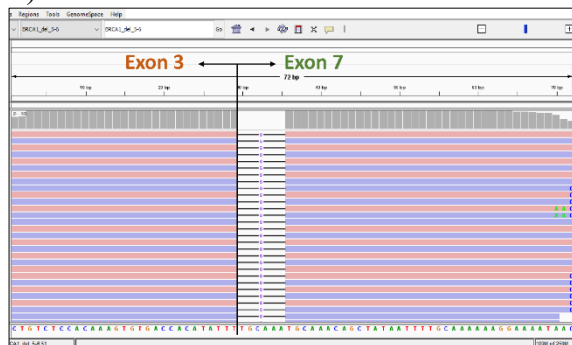
* U varianty Δ9_10 nelze rozlišit, zda je kombinována s kanonickým (e11) či alternativně zkráceným exonem 11 (e11q).

Nejčastěji zastoupenou sestřihovou událostí byl výpadek několika exonů v řadě (mCA – multikazetová delece), který byl samostatně identifikován v 32 případech a v 12 případech současně s delecí způsobenou volným sestřihovým místem (viz kapitola 1.2.2.5. *Alternativní sestřih na základě volných sestřihových míst*; +SDSΔ nebo +SASΔ; Obr. 28).

A)



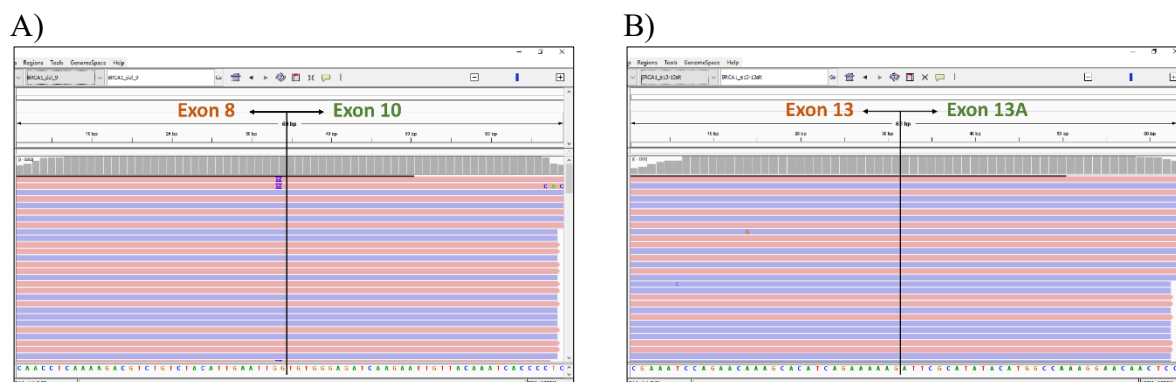
B)



Obrázek 28: Příklady nalezených multikazetových delecí. A) Nalezená multikazetové delece Δ3_5 **B)** multikazetová delece Δ5_6 ve spojení s volným sestřihovým místem 7p. Identifikovaná čtení jsou zobrazena v prohlížeči IGV.

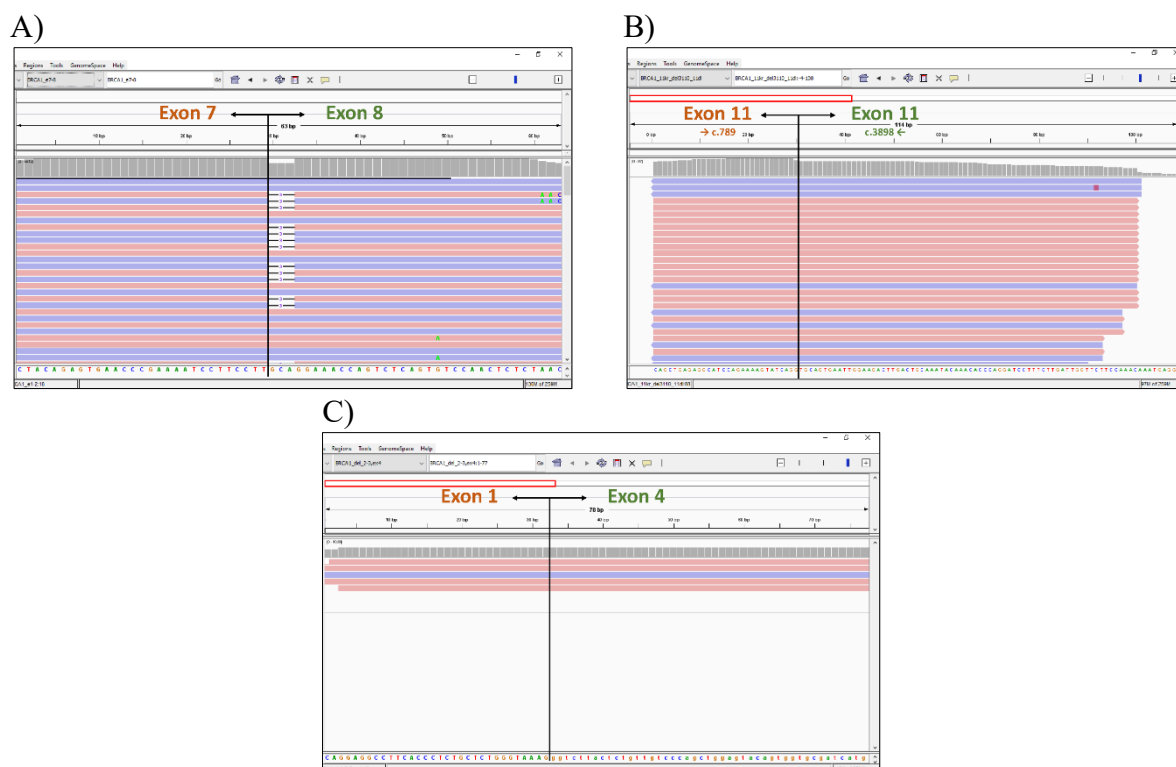
Dalším častým případem alternativního sestřihu byl výpadek jednoho exonu (CA – kazetová delece) nebo zahrnutí alternativního exonu či vnitřní části intronu do transkriptu (C▼ – kazetová inserce). Tyto varianty byly detekovány v 16 (CA) a 11 (C▼) případech osamoceně,

v jednom případě byla nalezena kombinace inserce a delece a v dalších 4 případech byly nalezeny jednotlivé inserce nebo delece v kombinaci s volným sestřihovým místem (Obr. 29).



Obrázek 29: Příklady nalezených kazetových variant. A) – delece exonu 9 ($\Delta 9$). B) – inserce alternativního exonu 13 ($\blacktriangledown 13A$) zobrazené v prohlížeči mapovaných čtení IGV.

Samostatné využití volných sestřihových míst bylo identifikováno ve 12 případech (Obr. 30A) a na závěr byly nalezeny tři případy intronizace vnitřní části exonu 11 (Obr. 30B) a jeden případ kombinace multikazetové delece (e2 a e3) s kazetovou insercí (alt. e4; Obr. 30C).

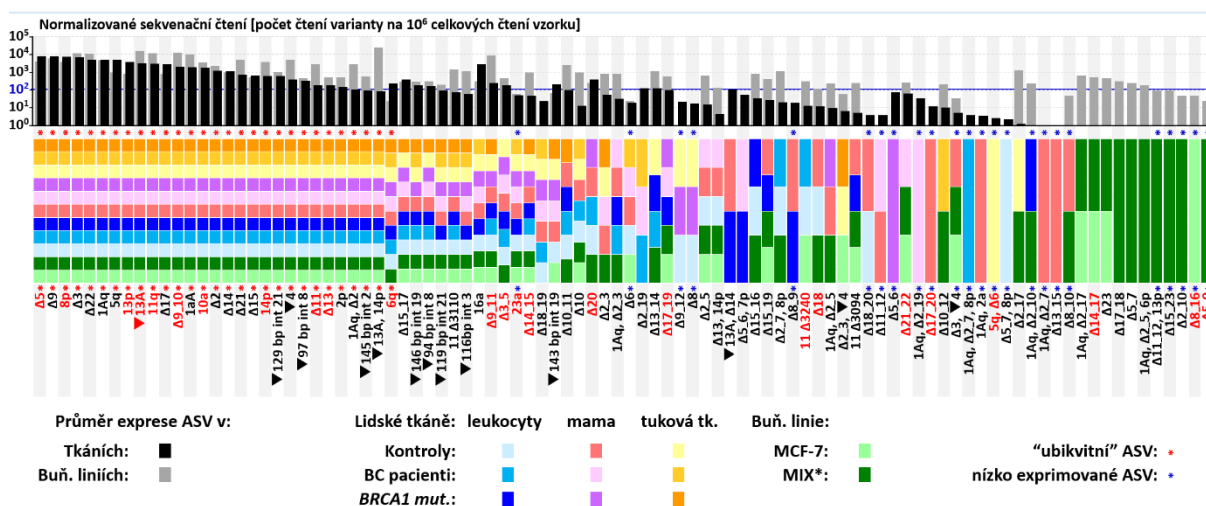


Obrázek 30: Unikátní případy alternativního sestřihu identifikovaného v transkriptu BRCA1. A) sestřih na základě využití volných sestřihových míst B) intronizace části exonu 11 (11 $\Delta 3110$). C) delece exonu 2 a 3 současně s insercí alternativního exonu 4 ($\Delta 2_3$, $\blacktriangledown 4$). Čtení jsou zobrazená v prohlížeči mapovaných čtení IGV.

4.2.6 Tkáňová specifičnost nalezených variant

V biologických vzorcích pacientů s karcinomem prsu a zdravých kontrol se nejširší spektrum exprimovaných ASV nacházelo v mamární tkáni (72). Leukocyty periferní krve exprimovaly 67 různých variant a přilehlá tuková tkáň 54 (viz Tab. 6). Ve všech vyšetřovaných cDNA syntetizovaných z lidského genetického materiálu bylo přítomno 29 variant, které jsou označeny jako ubikvitní / predominantní (Obr. 31). V testovaných buněčných liniích bylo nalezeno 76 ASV, z nichž 11 bylo pro dané linie unikátní (Obr. 31).

Nalezené kvalitativní a kvantitativní rozdíly mezi zdravými kontrolami, pacienty s karcinomem prsu a nosiči *BRCA1* mutace byly na úrovni semikvantitativního porovnání mapovaných výsledků malé (Tab. 6).



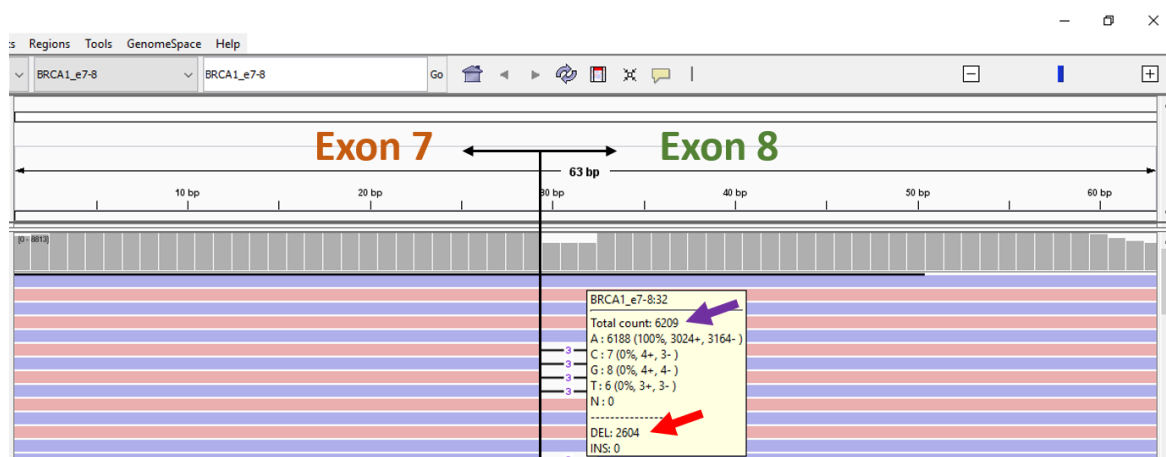
Obrázek 31: Semikvantitativní porovnání nalezených sestřihových variant. Černými a šedými sloupci je vyznačena četnost nalezené varianty normalizovaná na milion čtení. Červenými hvězdičkami je označena ubikvitní / predominantní sestřihová událost, modrými hvězdičkami sestřihová událost s nízkou četností. Barevným kódem je kvalitativně vyznačeno, ve kterých tkáních byla daná událost nalezena.

Proteinový produkt *BRCA1* může potencionálně vznikat pouze ze 25 identifikovaných variant, které zachovávají čtecí rámec („*in-frame*“). Jedenáct z těchto variant bylo označeno jako predominantní, tedy vyskytující se napříč vyšetřovanými soubory. Z těchto jedenácti variant je devět popsáno v předchozí literatuře ($\Delta 5$; 8p; $\Delta 9_{10}$; $\Delta 11$; 11q; 13p; $\nabla 13A$; $\nabla 13A,14p$; 14p) a dvě varianty byly identifikovány nově (6q; 10a). Další 4 „*in-frame*“ varianty ($\Delta 3_5$; $\Delta 9_{11}$; $\Delta 14_{15}$ a 23a) byly detekovány ve většině vyšetřovaných vzorků.

Zbýlých 10 variant ($\Delta 5_9$; 5q; $\Delta 6$; $\Delta 8_{16}$; 11 $\Delta 3240$; $\Delta 14_{17}$; $\Delta 17_{19}$; $\Delta 17_{20}$; $\Delta 18$; $\Delta 20$; $\Delta 21-22$) vykazovalo na základě identifikace variant pomocí NGS spíše slabší míru exprese.

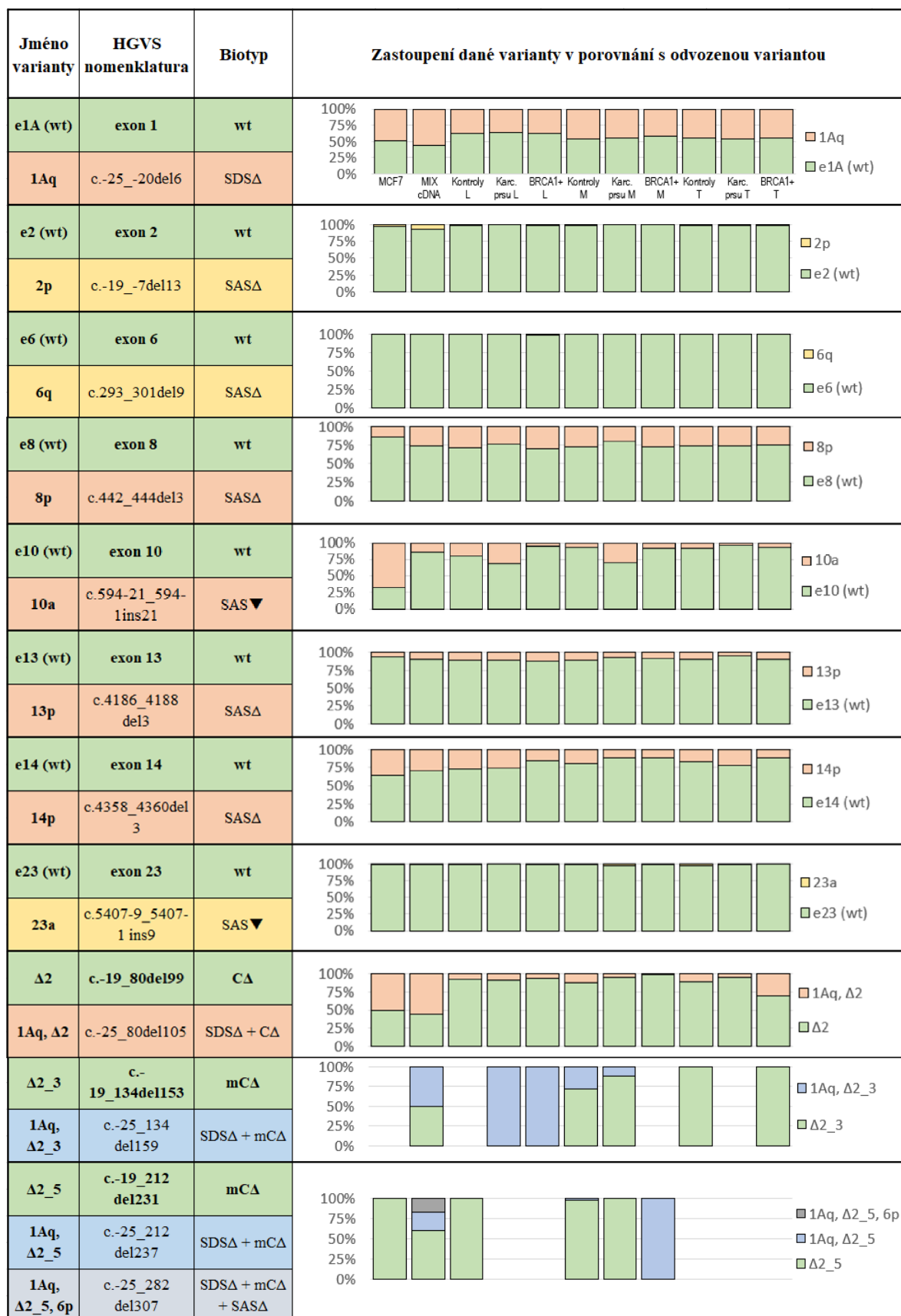
4.2.7 Kvantitativní posouzení variant s volnými sestřihovými místy

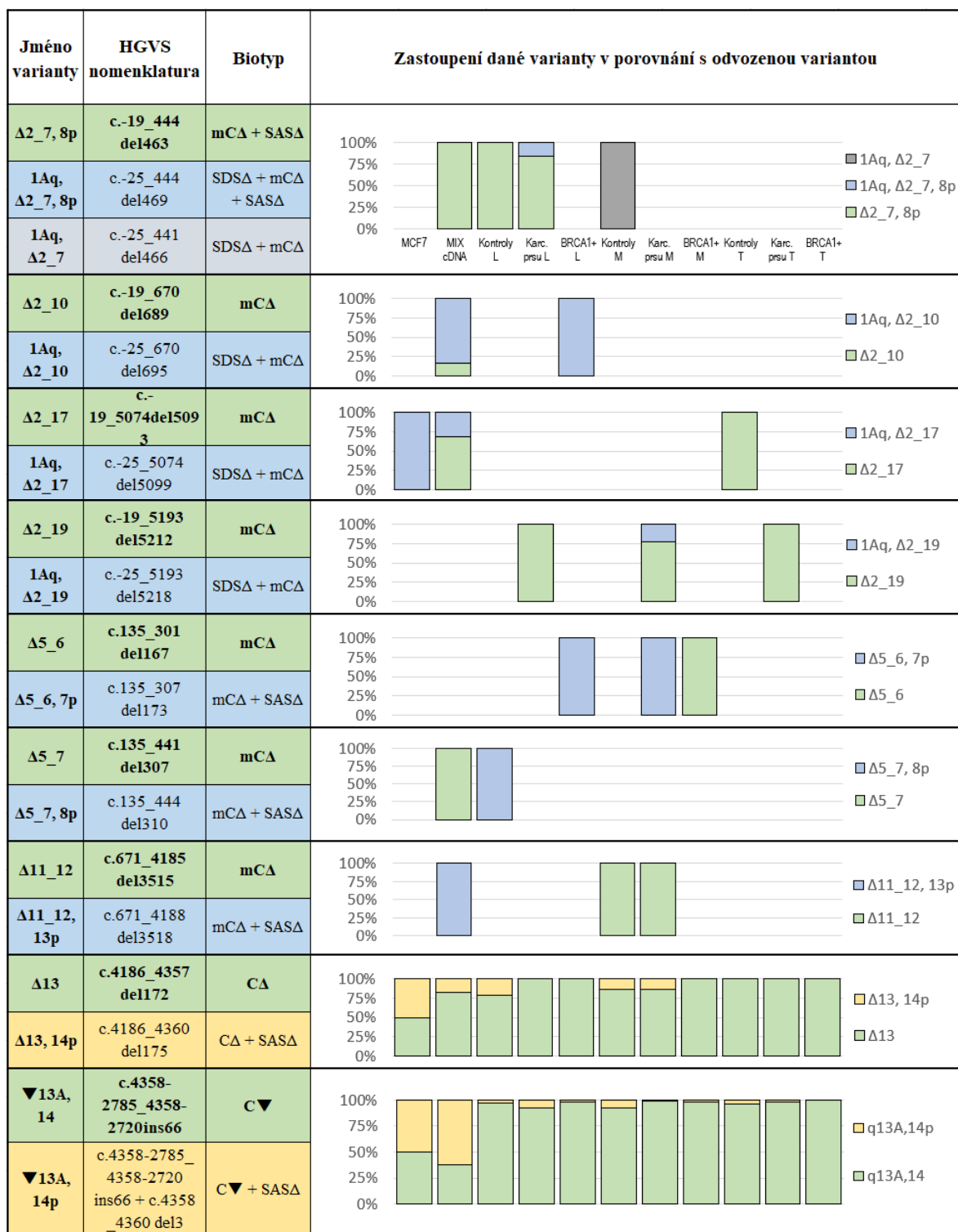
Použitý metodický přístup v identifikaci ASV nám umožňuje přímo kvantifikovat většinu variant lišících se na základě volných sestřihových míst bez další nutnosti kvantifikace pomocí RT-qPCR. Mapovací algoritmus čtení obsahující klasické sestřihové místo a čtení obsahující volné sestřihové místo ve většině případů namapoval na jednu sestřihovou událost. Pro kvantitativní posouzení bylo v prohlížeči mapovaných čtení potřeba odečíst poměr zastoupených frakcí (Obr. 32). V případech čtení mapovaných na jednu sestřihovou událost byl rozdíl v délkách jednotlivých ampliconů (s a bez využití volného sestřihového místa) minimální, a proto preferenční míra amplifikace kratších ampliconů zanedbatelná.



Obrázek 32: Odečtení poměru varianty s klasickým sestřihovým místem oproti variantě s volným sestřihovým místem (znázorněno na příkladu varianty 8p) v programu IGV. Tento program umožňuje zobrazení dodatečných informací k mapovaným čtením. Fialovou šipkou je označen celkový počet identifikovaných čtení v místě delece (tzn. čtení neobsahující deleci – v tomto případě kanonický sestřih), červenou šipkou je označeno množství čtení obsahující deleci (v tomto případě sestřih na základě volného sestřihového místa 8p).

Ze všech nalezených SDS a SAS variant (Splice donor site / Splice acceptor site; viz Tab. 6) bylo přímo kvantifikováno 22 případů s volným sestřihovým místem (Obr. 33; příloha 3):





Obrázek 33: Využití volných sestřihových míst. Zeleně – varianta od které je volná sestřihová varianta odvozena. **Červeně** – majoritně se vyskytující varianta. **Žlutě** – minoritně se vyskytující varianta. **Modře** – vzácně se vyskytující varianta. **Karc. prsu** – pacientky s karcinomem prsu **BRCA1+** – pacientky s mutací v *BRCA1*. **L** – leukocyty periferní krve; **M** – mamární tkáň, **T** – tuková tkáň.

Varianty 1Aq; 8p; 10a; 13p; 14p a 1Aq,Δ2; byly detekovány ve většině vzorků lidských tkání s četností > 5 % (v Obr. 33 a příloha 3 – červeně). Varianty 2p; 6q; 23a; Δ13,14p a ▼13A,14p se vyskytovaly napříč vzorky v míře nižší (v Obr. 33 a příloha 3 – žlutě). Kombinované varianty 1Aq,Δ2_3; 1Aq,Δ2_5; 1Aq,Δ2_5,6p; 1Aq,Δ2_7,8p; 1Aq,Δ2_7; 1Aq,Δ2_10; 1Aq,Δ2_17; 1Aq,Δ2_19; Δ5_6,7p; Δ5_7,8p; Δ11_12,13p byly detekovány pouze v některých případech (v Obr. 33 a příloha 3 – modře).

4.3 Kvantifikace vybraných ASV v biologických vzorcích pomocí RT-qPCR

4.3.1 Analyzovaný soubor vzorků

Po analýze hrubých RT-qPCR dat od 596 lidských vzorků, které prošly úspěšnou syntézou cDNA (viz kapitola 4.1.1 *Izolace a kontrola kvality celkové RNA z lidských tkání*) bylo z důvodu nízkého Cp referenčních oblastí BRCA1 mRNA ($C_p > 25$) vyřazeno 19 % naměřených singletů vzorků (tj. 269 z celkových 1408 singletů). Vyhodnocena byla finální data od 115–145 žen (v závislosti na vyšetřované tkáni; Tab. 7).

V rámci souboru zdravých kontrol byly úspěšně analyzovány a navzájem porovnány vzorky pravé a levé prsní tkáně u 40 (prsní) a 43 (tuková) žen.

Tabulka 7: Počet žen (vzorků) v jednotlivých skupinách vyšetřovaných souborů a tkání, které byly úspěšně analyzovány pomocí RT-qPCR a statisticky vyhodnoceny. L – leukocyty; M – mamární tkáň, T – tuková tkáň.

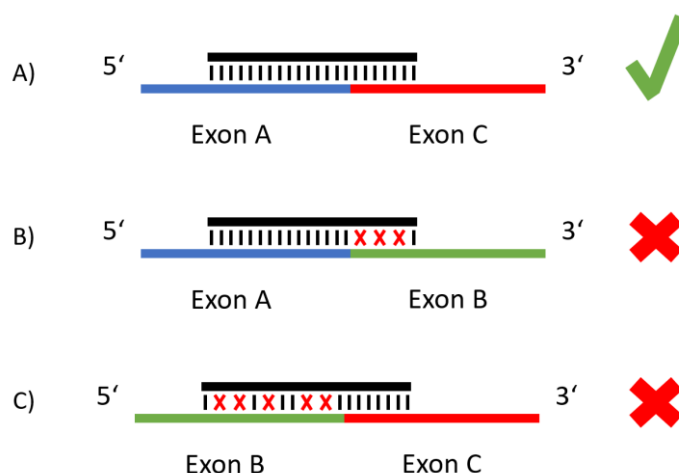
Název	BRCA1 mut.			BRCA1 wt		
	L	M	T	L	M	T
Ženy bez karcinomu prsu	70 % 7 / 10	70 % 7 / 10	70 % 7 / 10	51 % 39/76	72 % 55/76	72 % 55/76
Ženy s karcinomem prsu	29 % 2/7	57 % 4/7	43 % 3/7	60 % 74/123	64 % 79/123	54 % 67/123
Úspěšnost celkem	59,3 % 115-145 / 216					

Skupiny pacientů s mutací v *BRCA1* byly z návazných statistických analýz vyřazeny pro nedostatečné množství hodnotitelných vzorků.

4.3.2 Optimalizace qPCR reakce

Pro RT-qPCR s primery nasedajícími do unikátních exon-exonových spojení bylo kriticky důležité, aby se daný primer nevázal pouze na 5' nebo 3' exonovou část spoje. Pokud by se vázal na 5' část, vznikal by dlouhý kanonický PCR produkt (Obr. 34 B), který by se dal snadno odhalit pomocí kontrolní elektroforézy PCR produktů nebo tzv. high resolution melting (HRM)

analýzy, která následovala u všech vzorků po každém qPCR běhu. Pokud by se primer vázal pouze na krátkou 3' exonovou část unikátního přechodu (viz Obr. 34 C), mohl by vzniknout falešně pozitivní PCR produkt neodpovídající templátu. Všechny primery byly proto testovány pomocí PCR reakce jak s cDNA templátem (ve kterém se vyskytovala daná sestřihová varianta), tak s plazmidovým templátem obsahujícím wt sekvenci BRCA1 (viz kapitola 4.2.2. *Optimalizace primerů pro multiplexní PCR reakce*).



Obrázek 34: Schéma možného nasedání primeru do exon-exonových spojení. A) primer navržen do unikátního exon-exonového spojení exonů A/C. B) Stejný primer chybně se vázající do exon-exonového spojení A/B. Pokud primer vykazoval dostatečnou vazebnou schopnost v 5' exonu B, vznikl dlouhý PCR amplikon obsahující sekvenci exonů A, B a C. C) Stejný primer chybně se vázající do exon-exonového spojení B/C. Pokud primer vykazoval dostatečnou vazebnou schopnost v 3' exonu B, vznikl stejný PCR amplikon jako v případě A), obsahující sekvenci exonů A, C.

Primery byly hodnocené jako vyhovující, pokud tvořily PCR produkt s cDNA templátem (správnost sekvence byla ověřena pomocí Sangerova sekvenování) a netvořily žádný produkt s plazmidovým templátem, tedy byly specifické pro konkrétní unikátní exon-exonové spojení.

4.3.3 Zhodnocení vybraných referenčních genů

Žádný známý gen lidského organismu nevykazuje stejnou míru exprese svých transkriptů ve všech tkáních, avšak některé geny jsou exprimovány s relativně stejnou mírou a pro potřeby RT-qPCR se dají použít jako geny referenční (RG).

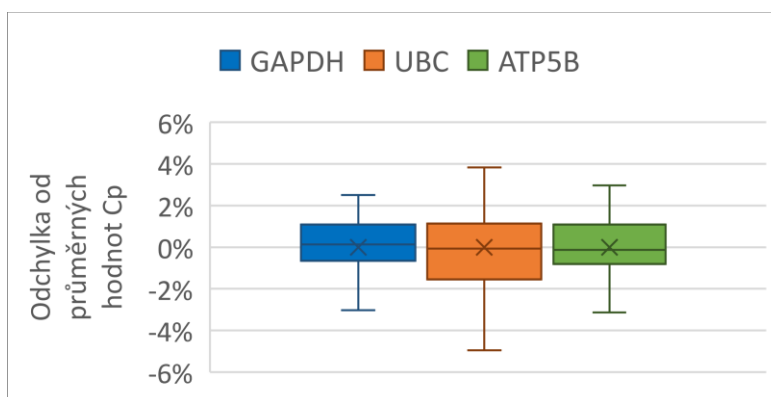
Pro kvantitativní posouzení exprese ASV BRCA1 ve vyšetřovaných tkáních byly vybrány, na základě předchozích zkušeností z kvantitativních analýz (Kleiblová et al., 2010) a výsledků

volně dostupných externích dat z expresní analýzy lidského transkriptomu (<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB2445/>), referenční geny *GAPDH*, *UBC* a *ATP5B* (Tab. 8).

Tabulka 8: Míry exprese referenčních genů ve vyšetřovaných tkáních. Data jsou čerpána z projektu Illumina bodyMap2 transcriptome (<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB2445/>). RPKM – normalizovaná hodnota množství čtení na kilobázi transkriptu na milion čtení (reads per kilobase milion).

Název referenčního genu	Míra exprese v leukocytech (RPKM)	Míra exprese v prsní tkáni (RPKM)	Míra průměrné exprese v tukové tkáni (RPKM)
<i>GAPDH</i>	378,46	435,31	345,29
<i>UBC</i>	810,76	617,48	631,75
<i>ATP5B</i>	142,08	128,90	77,50

Před samotnými analýzami naměřených dat byla posouzena přesnost práce a reprodukovatelnosti výsledků celkového experimentu. Nezávisle naměřená hrubá data ve formě Cp referenčních genů byla otestována ve výběru 50 nejkvalitnějších vzorků ze souboru (RIN > 8; vstup do syntézy cDNA = 800 ng). Experiment prokázal nesignifikantní rozptyly hodnot Cp jednotlivých referenčních genů v jednotlivých tkáních (Kruskal-Wallisův test pro porovnání více výběrů, $p=0,545$). Maximální odchylka od průměrné hodnoty u *UBC* byla 5 %. U *GAPDH* a *ATP5B* pak 3 % (Obr. 35).



Obrázek 35: Porovnání odchylky Cp referenčních genů od průměrných hodnot v souborech vyšetřovaných tkání. Pro každou vyšetřovanou tkáň byla vypočítána průměrná hodnota Cp u daného referenčního genu. Dále byly kalkulovány jednotlivé odchylky od této hodnoty. Odchylky v jednotlivých tkáních jsou znázorněny společně pro každý RG.

Expresí RG v jednotlivých tkáních byla následně porovnána s expresemi naměřenými v projektu Illumina bodyMap2 (viz Tab. 8). Data z našeho i externího souboru byla normalizována na míru exprese *GAPDH* v prsní tkáni a poměrově porovnána (Tab. 9).

Tabulka 9: Rozdíly v expresi Illumina bodyMap 2 oproti expresi z RT-qPCR (data jsou normalizována na expresi *GAPDH* v prsní tkáni – 0)

Název referenčního genu	% rozdíl exprese v leukocytech	% rozdíl exprese prsní tkáni	% rozdíl exprese v tukové tkáni
<i>GAPDH</i>	-1,42	0,00	1,87
<i>UBC</i>	103,36	26,06	-12,10
<i>ATP5B</i>	127,40	-10,62	-62,90

Z pohledu absolutní kvantifikace nejlépe korelovaly exprese *GAPDH*, které jsou ve vyšetřovaných tkáních srovnatelné se získanými daty Illumina bodyMap2. Míry expresí *UBC* a *ATP5B* se s porovnávanými daty lišily, obzvláště v leukocytech periferní krve. Vzhledem k rozsáhlosti našeho souboru (přes 100 osob vs. 1 osoba v případě Illumina bodyMap2) a konzistence námi naměřených hrubých hodnot Cp (Obr. 35) jsme žádný RG z dalších analýz nevyřadili.

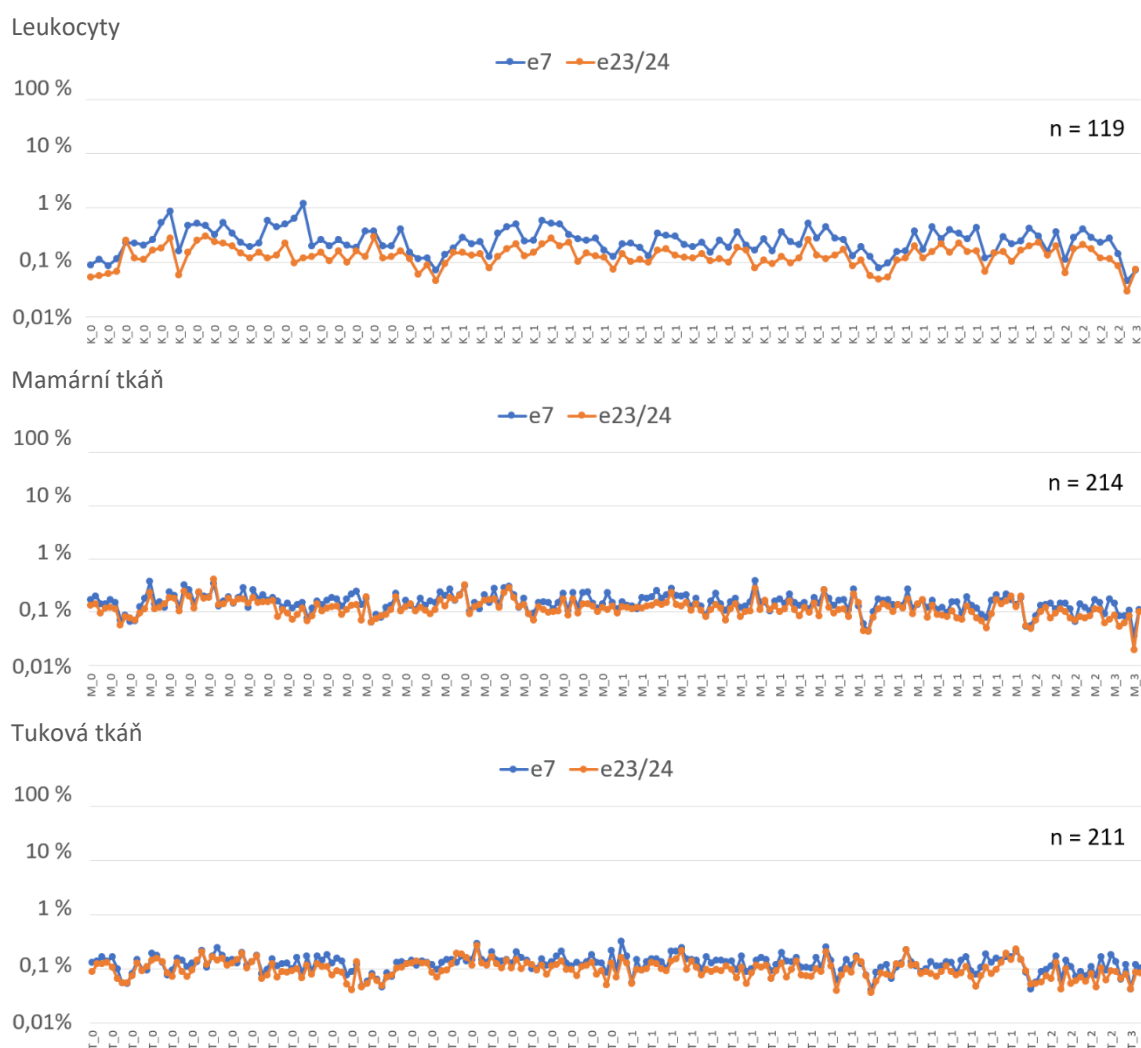
4.3.4 Celková exprese BRCA1 ve vyšetřovaných tkáních

Celková míra exprese všech transkriptů BRCA1 byla měřena pomocí dvou amplikonů lokalizovaných v různých částech referenčního transkriptu. Prvním byl amplikon zasahující do přechodu exonů 23 a 24 (viz příloha 2), který byl vybrán z důvodu absence sestřihových variant postihujících tyto exony ve vyšetřovaných lidských tkáních³ (viz. Tab. 6). Druhý amplikon byl lokalizován do exonu 7, který zohledňuje variantu BRCA1 IRIS (či případně jinou variantu

³ Jedinou dosud známou výjimkou, nepřepisující 3' koncovou část genu do transkriptu, je alternativní genový produkt BRCA1-IRIS (viz kapitola 1.3.3 *Alternativní sestřihové varianty BRCA1 a jejich význam*). U tohoto transkriptu je ukončena transkripce v intronu 11 a proto naměřené exprese amplikonu e23/24 nezahrnují expresi této unikátní varianty (ani případných jiných, podobných variant).

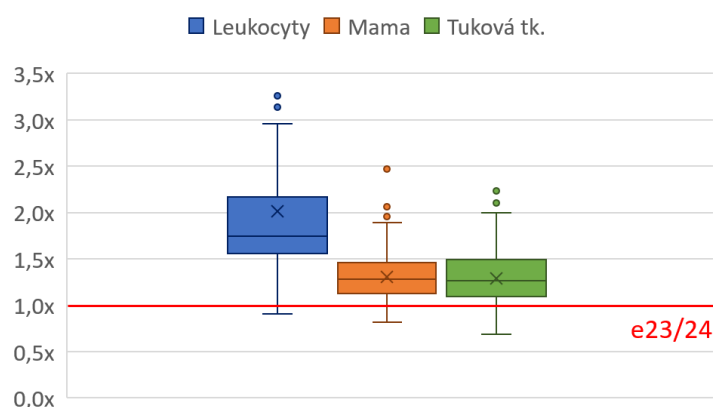
s ukončenou transkripcí za exonem 7). Zároveň nebyla identifikována žádná významná sestřihová varianta postrádající exon 7 (v součtu bylo identifikováno pouze 1643 čtení různých minoritních sestřihových událostí postrádající exon 7, viz Tab. 6).

Oba referenční amplikony reprezentující celkovou expresi BRCA1 vykazují v porovnání se všemi referenčními geny velmi nízkou (100x až 1000x nižší) míru exprese napříč všemi vyšetřovanými vzorky (Obr. 36).



Obrázek 36: Porovnání míry exprese referenčních oblastí BRCA1 normalizovaných na expresi vybraných referenčních genů. Modrými (e7) a oranžovými (e23/24) body jsou označeny relativní exprese daných variant u jednotlivých vzorků. Písmenem (K – leukocyty periferní krve; M – mamární tkáň; T – tuková tkáň) a číslem (0 = zdravé kontroly, 1 = pacientky s karcinomem prsu, 2 = zdravé a *BRCA1* mutované a 3 = s karcinomem prsu a *BRCA1* mutované) jsou označeny jednotlivé soubory. n - počet vyšetřovaných vzorků v dané tkáni.

Z grafu je patrný trend vyšší exprese amplikonu e7 oproti amplikonu e23/24. Pro posouzení rozdílu exprese e7 a e23/24 byla souhrnná data v leukocytech, prsní a tukové tkáni graficky porovnána (Obr. 37).



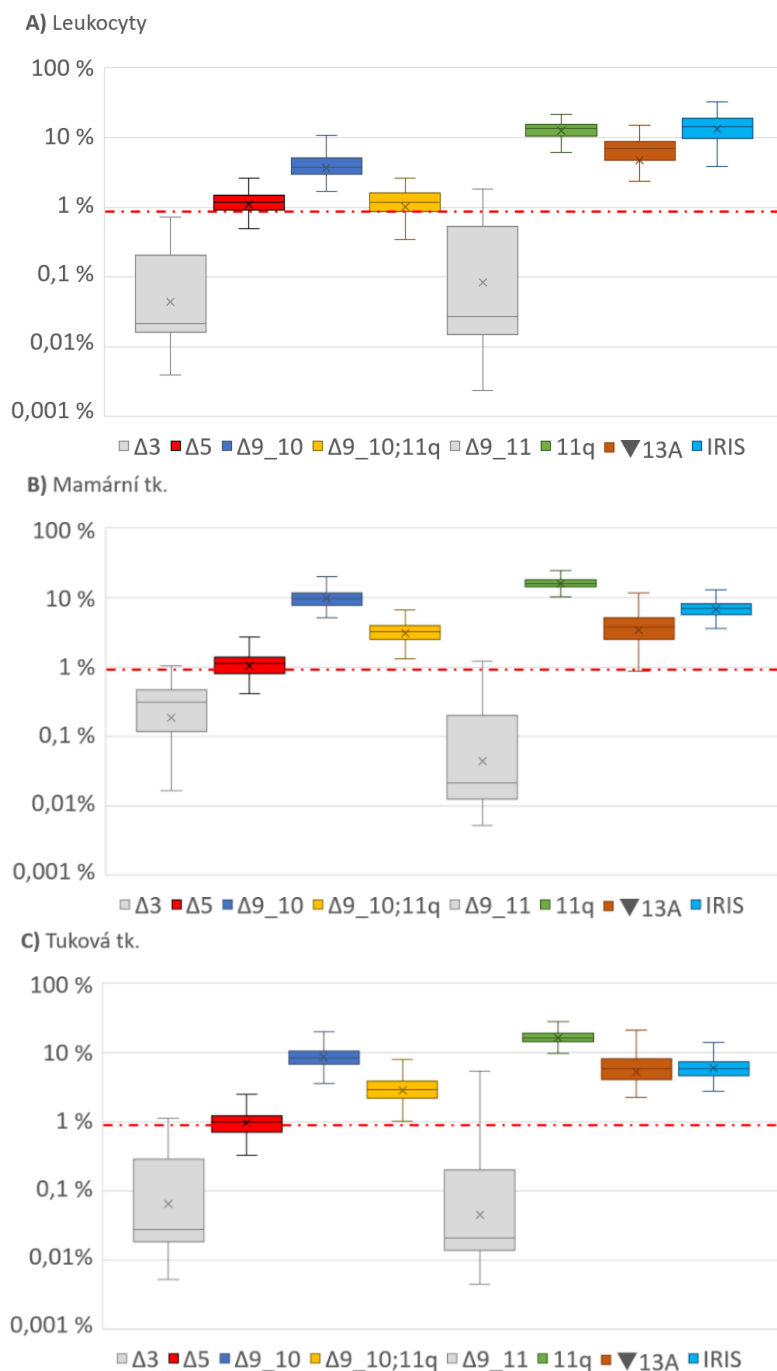
Obrázek 37: Porovnání exprese BRCA1 e7 proti expresi BRCA1 e23/24 ve všech vzorcích napříč vyšetřovanými tkáněmi. Červenou linkou je znázorněna normalizovaná exprese e23/24 (1x).

Vyšší relativní exprese e7 proti exonu 23/24 prokazuje přítomnost jedné či více variant, které postrádají exony 23 a 24 (např. zmíněná BRCA1 IRIS, či další). Zároveň je patrná různá tkáňově specifická míra exprese e7, která naznačuje kvantitativně odlišné složení ASV BRCA1 v leukocytech oproti vyšetřovaným solidním tkáním. Z tohoto důvodu byla exprese jednotlivých ASV BRCA1 (včetně BRCA1 IRIS) porovnávána proti expresi BRCA1 e7.

4.3.5 Relativní exprese ASV BRCA1

Z celkových 94 identifikovaných variant pomocí NGS bylo přímo kvantifikováno 22 variant z mapovaných dat. U zbývajících 72 variant nebyla přesně kvantifikována míra jejich exprese, protože nebylo možné reflektovat a zpětně spočítat rozdílnou efektivitu jednotlivých mPCR reakcí či ztrátu různě dlouhých produktů mPCR v promývacích krocích. Kvantifikace všech 72 variant pomocí přesné RT-qPCR metody byla kvůli omezenému biologickému materiálu nemožná, a proto byly pro další analýzu vybrány pouze varianty predominantní, zachovávající čtecí rámec – potencionálně vedoucí ke vzniku proteinového produktu ($\Delta 3$; $\Delta 5$; $\Delta 9_{10}$ s navazující krátkou i dlouhou formou exonu 11; 11q; $\Delta 11$; $\nabla 13A$). Do souboru byly dále přidány „*in-frame*“ varianty, které byly detekovány alespoň v 6 z 9 lidských tkání ve stabilní míře ($\Delta 3_5$ a $\Delta 9_{11}$) a popsaná varianta BRCA1 IRIS. Po provedení RT-qPCR s primery navrženými do unikátních exon-exonových přechodů na tomto souboru mohla být dále

statisticky vyhodnocena relativní exprese většiny variant (Obr. 38). Varianty $\Delta 3_5$ a $\Delta 11$ nebylo možné hodnotit z důvodu hrubých hodnot C_p za detekčním limitem metody⁴ ($C_p > 33$) u většiny vzorků, i přes preamplifikaci cDNA.



Obrázek 38: Relativní exprese ASV BRCA1 ve vyšetřovaných tkáních vztažena k celkové BRCA1 e7. Přerušovanou červenou linkou je vyznačen spolehlivý detekční limit metody pro kvantifikaci ASV BRCA1 (1 % z celkové BRCA1).

⁴ Detekční limit metody byl charakterizován v pilotních expresních pokusech v práci Hojný, 2012.

Varianty $\Delta 3$ a $\Delta 9_{-11}$ se exprimují ve všech tkáních ve velmi malé míře. Z rozptylu výsledků a hrubých hodnot C_p (31-33) je patrné, že míra jejich exprese naráží na detekční limit metody, proto také nebyly tyto varianty dále hodnoceny.

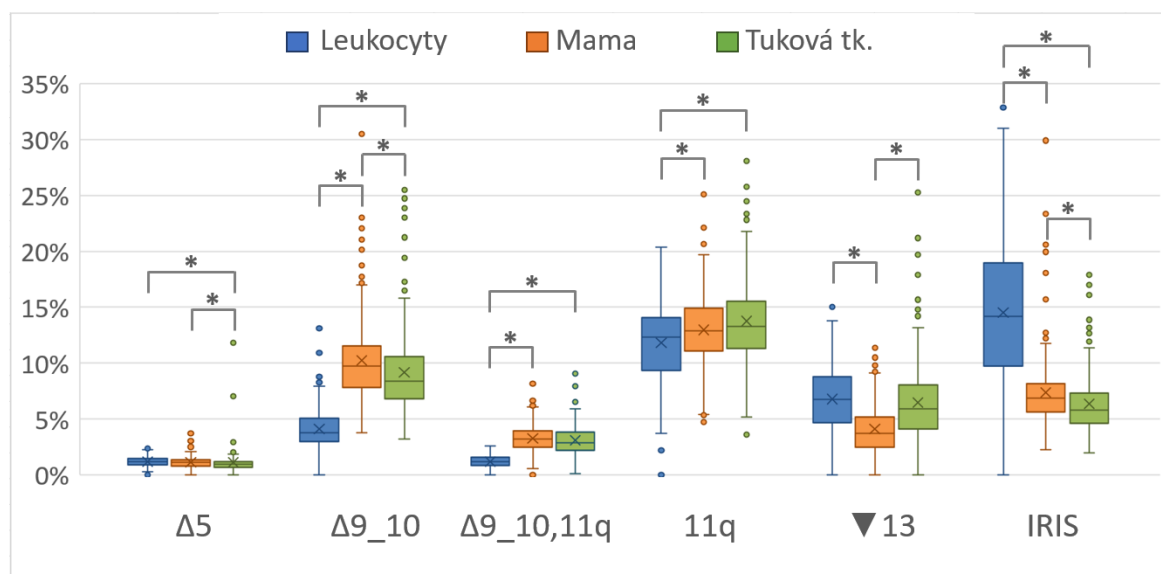
4.3.5.1 Tkáňově specifická exprese ASV BRCA1

Pro finální zhodnocení míry měřených ASV bylo nutné zohlednit celkovou expresi varianty 11q (která v experimentu díky návrhu primerů reprezentuje všechny případy krátké formy exonu 11 – i s delecí $\Delta 9_{-10}$). Exprese samotné varianty 11q byla získána odečtením exprese varianty $\Delta 9_{-10,11q}$. Po tomto přepočtu není součet měřených ASV BRCA1 v porovnání s celkovou BRCA1 uměle nadhodnocený. Hodnoty relativní exprese vyšetřovaných variant ve vyšetřovaných tkáních jsou shrnuty v Tab. 10 a graficky znázorněny na Obr. 39 a 40.

Tabulka 10: Relativní míra exprese vyšetřovaných variant BRCA1. Exprese jednotlivých variant je vyjádřena v % celkové exprese BRCA1 (e7). SD = směrodatná odchylka; SEM = střední chyba průměru. Pro statistické porovnání tkáňových skupin byl použit neparametrický Mann-Whitney test. P hodnoty jsou označeny: **červeně** $p > 0,0085$ a **zeleně** $p < 0,0081$. Hodnota $p < 0,0085$ byla použita u důvodu eliminace chyby typu 1. Pozn.: tyto výsledky nezohledňují míru exprese celkové BRCA1, která je v jednotlivých tkáních odlišná, viz Obr. 36).

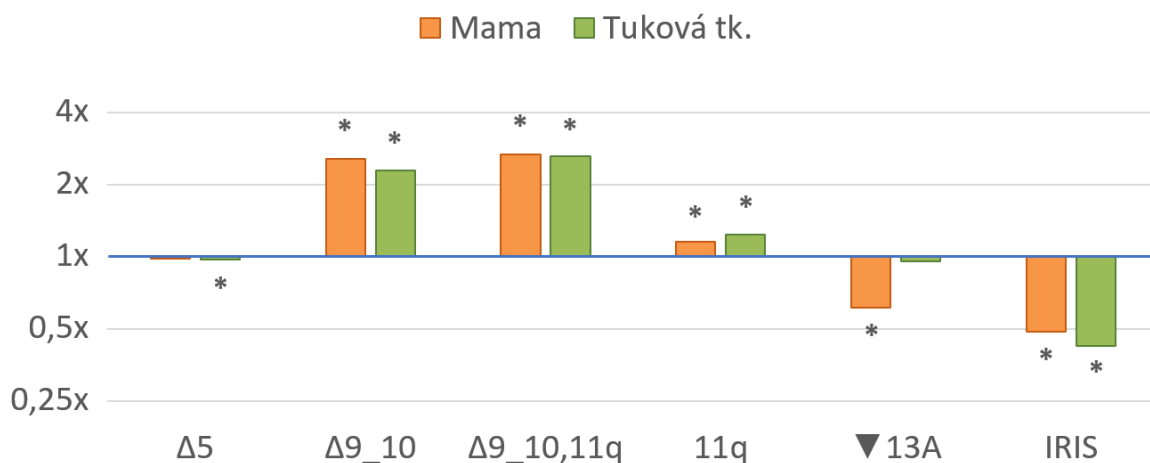
Název		$\Delta 5$	$\Delta 9_{-10};$ 11	$\Delta 9_{-10};$ 11q	11q	▼13A	IRIS
Leukocyty (L)	Exprese	1,13%	3,96%	1,22%	11,17%	6,70%	14,62%
	SD	0,44%	1,97%	0,55%	3,24%	3,14%	12,01%
	SEM	0,04%	0,18%	0,05%	0,29%	0,28%	1,09%
Mama (M)	Exprese	1,12%	10,21%	3,27%	12,95%	4,14%	7,16%
	SD	0,49%	3,67%	1,16%	2,93%	2,12%	2,75%
	SEM	0,03%	0,26%	0,08%	0,21%	0,15%	0,20%
	Poměr M vs. L	0,996	2,575	2,677	1,160	0,617	0,490
	P	0,5125	0	0	0,0000	0,0000	0
Tuková tk. (T)	Exprese	1,10%	9,11%	3,21%	13,91%	6,44%	6,24%
	SD	1,07%	3,37%	1,33%	3,19%	3,32%	2,11%
	SEM	0,08%	0,25%	0,10%	0,24%	0,25%	0,16%
	Poměr T vs. L	0,976	2,297	2,631	1,245	0,960	0,427
	P	0,0009	0	0	0,0000	0,1870	0
	Poměr T vs. M	0,980	0,892	0,983	1,074	1,555	0,872
	P	0,0018	0,0000	0,2330	0,0285	0,0000	0,0001

Varianty $\Delta 9_10$, $\Delta 9_10,11q$, $\nabla 13$ a IRIS vykazují odlišnou, tkáňově specifickou míru exprese vztažené k celkové BRCA1 (Obr. 39).



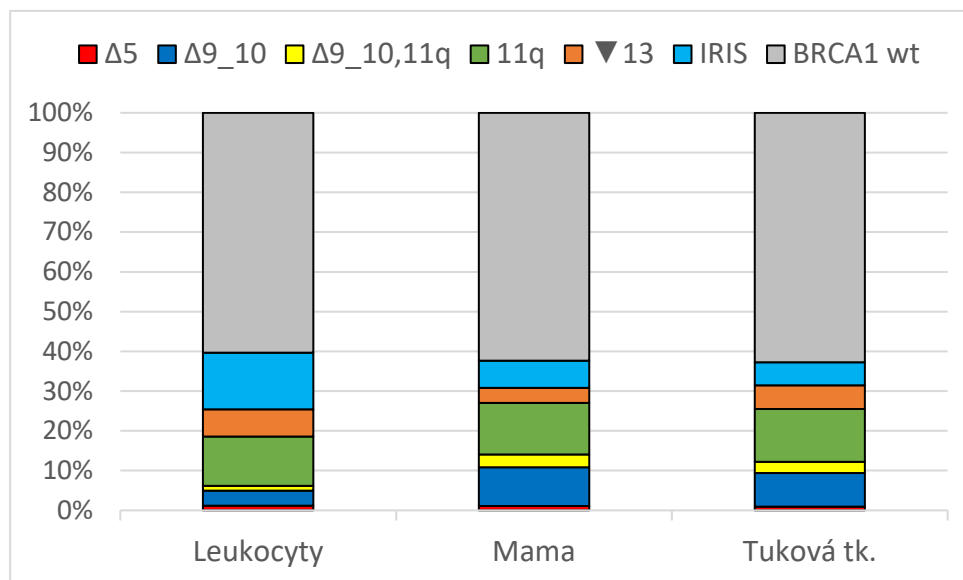
Obrázek 39: Graf míry exprese jednotlivých sestřihových variant v porovnání s celkovou expresí BRCA1 (100 %). Výsledky jsou zobrazeny ve formě krabicových grafů. Hvězdička značí statisticky signifikantní rozdíl exprese porovnávaných souborů.

Pro porovnání rozdílů průměrné exprese ve vyšetřovaných tkáních byla exprese variant mamární a přilehlé tukové tkáně vztažena k expresi variant kvantifikovaných v leukocytech (Obr. 40). Z grafu je patrný minimální rozdíl v expresích variant $\Delta 5$ a 11q, a to i přes statistickou signifikanci rozdílů většiny porovnávaných souborů.



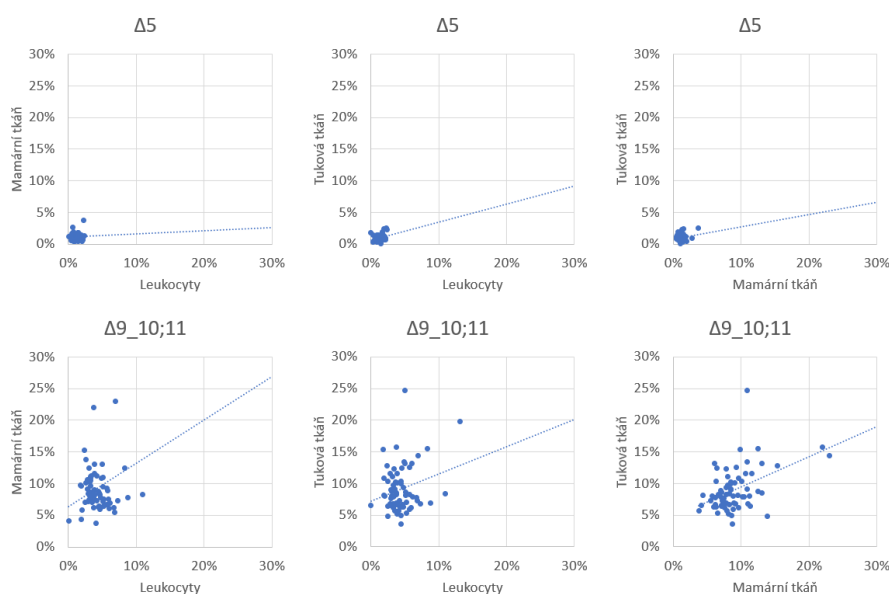
Obrázek 40: Rozdíly průměrných expresí jednotlivých ASV v mamární a tukové tkáni v porovnání s expresí v leukocytech. Modrá linka reprezentuje expresi variant v leukocytech (1x). Hvězdička značí statisticky signifikantní rozdíl exprese porovnávaných souborů.

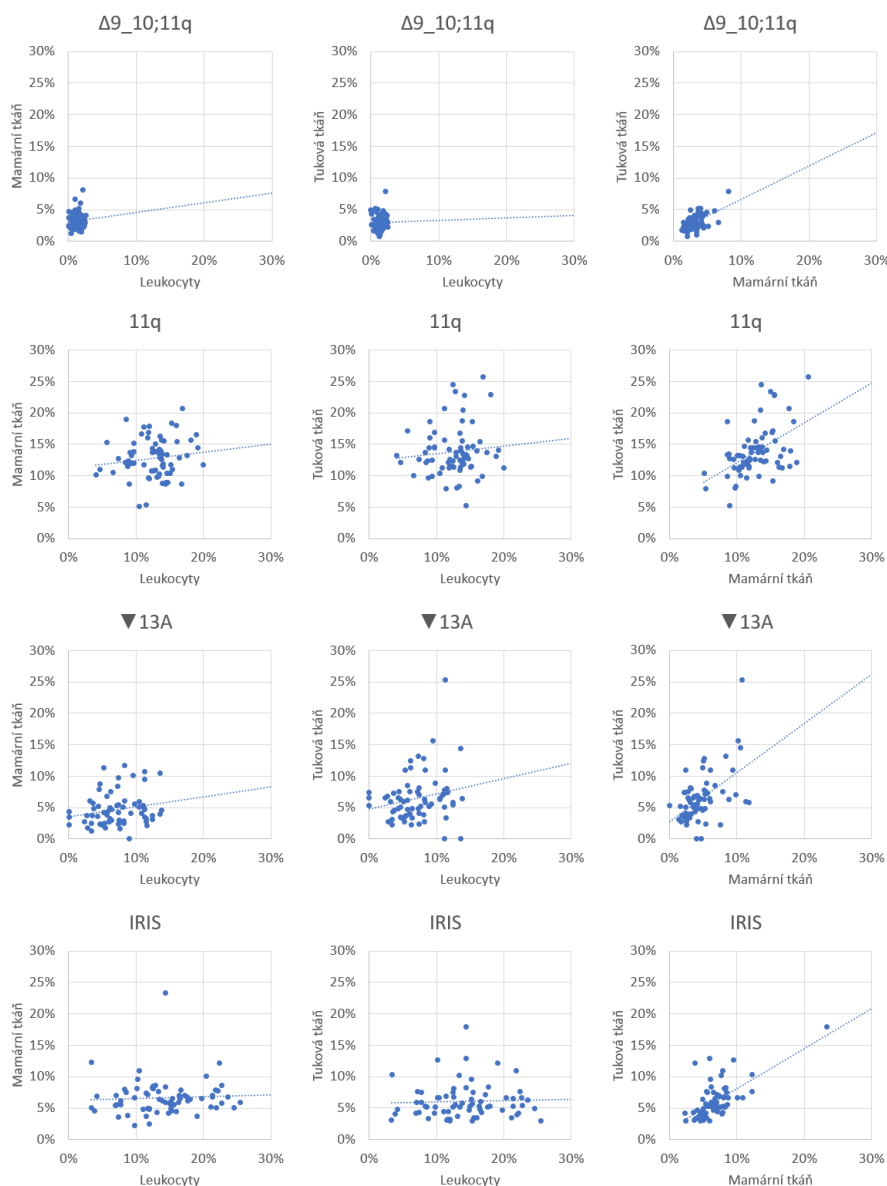
Souhrnná míra exprese vyšetřovaných variant z celkové BRCA1 (e7) dosahuje 39,7 % v leukocytech; 37,9 % v prsní tkáni a 37,3 % v tukové tkáni. Ve vyšetřovaných tkáních je souhrnná exprese těchto ASV srovnatelná, avšak míra exprese jednotlivých ASV je tkáňově specifická (Obr. 41).



Obrázek 41: Souhrnná exprese měřených ASV BRCA1 v jednotlivých tkáních. Barevně jsou vyznačeny podíly jednotlivých variant z celkové BRCA1, šedě je vyznačena exprese odpovídající BRCA1 wt.

Na neselektovaném souboru 70 žen (od kterých byla k dispozici expresní data pro každou variantu ve všech třech vyšetřovaných tkání) byla provedena korelace expresí jednotlivých variant mezi tkáněmi (Obr. 42).





Obrázek 42: Grafy korelace expresí jednotlivých variant ve vyšetřovaných tkáních. Modré body znázorňují jednotlivá měření expresí dané varianty v daných tkáních, které jsou vztažena k celkové expresi BRCA1 (e7). Tečkovaná modrá přímka znázorňuje spojnici trendu grafů.

Expresie jednotlivých variant v leukocytech periferní krve většinou individuálně nekoreluje (statisticky signifikantně) s expresí příslušných variant v mamární a přilehlé tukové tkáni. Výjimkou je korelace exprese varianty $\Delta 9_10$ s dlouhou formou exonu 11 ve všech tkáních a varianty $\Delta 5$ v leukocytech a mamární tkáni (tato varianta nicméně vykazuje velmi nízkou relativní expresi). Korelační faktory jsou však v těchto případech nízké ($< 0,5$).

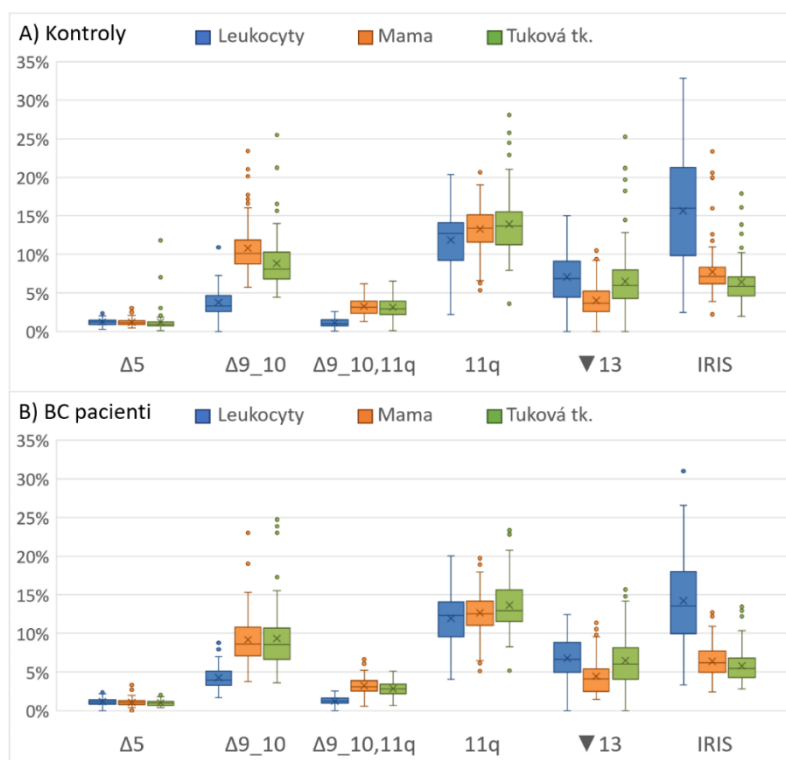
Naopak exprese jednotlivých variant v mamární tkáni statisticky signifikantně koreluje s expresí stejných variant ve tkáni tukové s výjimkou málo exprimované varianty $\Delta 5$ (Tab. 11) a zároveň jsou ve většině případů korelační faktory $> 0,5$.

Tabulka 11: Výsledky korelace jednotlivých variant ve vyšetřovaných tkáních. Hodnoty odpovídají příslušnému korelačnímu faktoru. Červeně jsou vyznačeny korelační faktory se statistickou signifikancí ($p < 0,05$)

Název varianty	$\Delta 5$	$\Delta 9_10$; 11	$\Delta 9_10$; 11q	11q	▼13A	IRIS
Leukocyty vs. Mama	0,046	0,335	0,077	0,138	0,205	0,055
Leukocyty vs. Tuková tk.	0,295	0,244	0,021	0,101	0,210	0,042
Mama vs. Tuková tk.	0,217	0,551	0,555	0,524	0,523	0,653

4.3.5.2 Relativní exprese ASV BRCA1 ve skupině zdravých kontrol a pacientek s karcinomem prsu

Z výsledků popisovaných v předchozí kapitole je patrný velký rozptyl exprese některých variant ($\Delta 9$ -10, 11q, ▼13 a IRIS), který může být způsoben odlišnou mírou exprese těchto variant v závislosti na vyšetřované podskupině. Proto bylo provedeno porovnání expresí zvlášť u souboru zdravých kontrol (76 žen) a pacientek s karcinomem prsu (123 žen; Obr. 43).



Obrázek 43: Grafy míry exprese jednotlivých sestřihových variant u skupiny zdravých kontrol (A) a pacientek s karcinomem prsu (B) v porovnání s celkovou expresí BRCA1 (100 %).

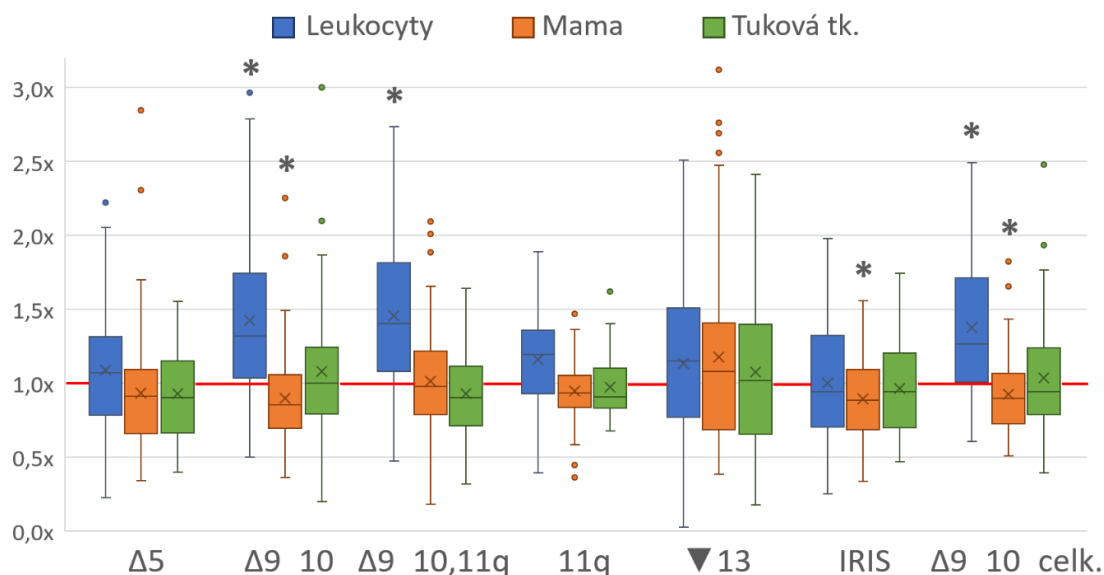
Z grafu je patrné, že rozptyly hodnot nesouvisí s diagnózou karcinomu prsu, ale jsou přirozené pro oba vyšetřované soubory. Naměřená data obou podskupin žen byla dále statisticky vyhodnocena a analýza odhalila několik případů signifikantně se lišící míry exprese u souboru žen s karcinomem prsu oproti zdravým kontrolám (Tab. 12).

Tabulka 12: Statistické porovnání exprese jednotlivých variant souboru pacientek s karcinomem prsu proti zdravým kontrolám. Poměr exprese vyjadřuje násobek exprese u skupiny pacientek s karcinomem prsu proti zdravým kontrolám. Expresce jednotlivých variant je vyjádřena v % celkové exprese BRCA1 (e7). SD = směrodatná odchylka; SEM = střední chyba průměru. Pro statistické porovnání byl použit neparametrický Mann-Whitney test. Vypočítané P hodnoty jsou označeny: **červeně** $p > 0,0073$ a **zeleně** $p < 0,0073$. Hodnota $p < 0,0073$ byla použita z důvodu eliminace chyby typu 1. Pozn.: tyto výsledky nezohledňují míru exprese celkové BRCA1, která je v jednotlivých tkáních odlišná, viz Obr. 36).

Název			$\Delta 5$	$\Delta 9_{10}; 11$	$\Delta 9_{10}; 11q$	11q	▼13A	IRIS	$\Delta 9_{10}$ celkem *
Leukocyty	Kontroly	Expresce	1,07%	3,26%	1,04%	10,45%	6,32%	17,69%	4,31%
		SD	0,41%	1,88%	0,64%	3,63%	3,40%	19,46%	2,09%
		SEM	0,07%	0,30%	0,10%	0,58%	0,54%	3,12%	0,33%
	Pac. s karc. prsu	Expresce	1,12%	4,24%	1,31%	11,59%	6,91%	13,26%	5,55%
		SD	0,41%	1,67%	0,48%	2,97%	3,09%	5,26%	1,84%
		SEM	0,05%	0,19%	0,06%	0,35%	0,36%	0,61%	0,21%
	Poměr exprese		1,047	1,298	1,259	1,109	1,093	0,750	1,288
	P		0,6703	0,0012	0,0048	0,1115	0,2450	0,7375	0,0006
Mama	Kontroly	Expresce	1,18%	10,78%	3,25%	13,24%	4,04%	7,50%	14,04%
		SD	0,46%	3,27%	1,10%	2,70%	1,96%	2,61%	3,94%
		SEM	0,05%	0,32%	0,11%	0,26%	0,19%	0,26%	0,39%
	Pac. s karc. prsu	Expresce	1,05%	9,18%	3,22%	12,72%	4,44%	6,34%	12,40%
		SD	0,45%	3,03%	1,11%	2,77%	2,36%	1,86%	3,42%
		SEM	0,05%	0,34%	0,12%	0,31%	0,27%	0,21%	0,39%
	Poměr exprese		0,893	0,851	0,990	0,960	1,099	0,846	0,883
	P		0,0317	0,0002	0,8381	0,1357	0,4527	0,0011	0,0047
Tuková tk.	Kontroly	Expresce	1,24%	8,99%	3,26%	14,19%	6,42%	6,27%	12,25%
		SD	1,39%	3,09%	1,16%	3,23%	3,40%	2,01%	3,77%
		SEM	0,14%	0,31%	0,12%	0,32%	0,34%	0,20%	0,38%
	Pac. s karc. prsu	Expresce	0,90%	8,89%	2,87%	13,52%	6,57%	5,77%	11,76%
		SD	0,28%	3,52%	0,94%	2,72%	3,10%	1,96%	4,03%
		SEM	0,03%	0,44%	0,12%	0,34%	0,39%	0,24%	0,50%
	Poměr exprese		0,725	0,989	0,880	0,953	1,024	0,921	0,960
	P		0,0579	0,5021	0,0619	0,1661	0,7250	0,1600	0,2031

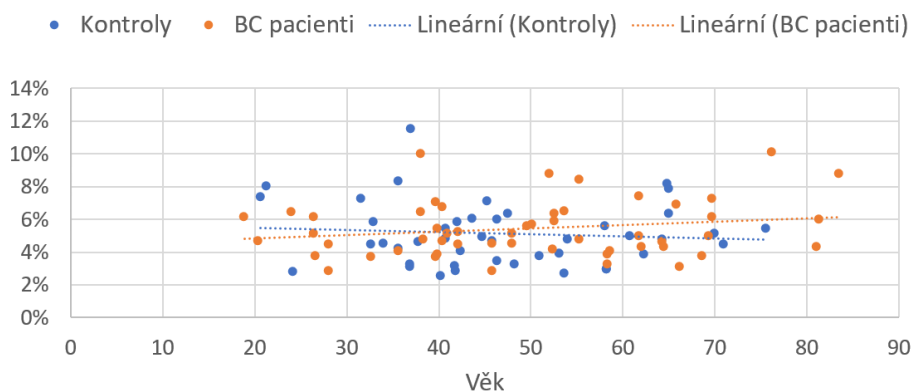
* Obě varianty s výpadkem exonů 9 a 10 byly sjednoceny a přidány do analýzy jako společná varianta, pro odstranění vlivu dlouhé či krátké formy exonu 11 a pro potvrzení signifikance dílčích výsledků.

Pro vytvoření grafického znázornění rozdílů porovnávaných souborů byla míra exprese ASV BRCA1 u pacientek s karcinomem prsu vztažena na míru exprese variant u zdravých kontrol (Obr. 44).



Obrázek 44: Graf rozdílů exprese jednotlivých ASV v souboru pacientek s karcinomem prsu vztažených na míru exprese zdravých kontrol. Červeně – exprese zdravých kontrol – 1x. Barevně jsou vyznačeny jednotlivé varianty. Hvězdičky = signifikantně se lišící ($p < 0,0073$).

Varianty s výpadkem exonů 9 a 10 v leukocytech periferní krve jsou jediné varianty, u kterých se signifikantně liší míra jejich exprese o více než 20 %. Průměrný věk vyšetřovaných skupin je však odlišný (zdravé ženy – 41,1 let; pacientky s karcinomem prsu – 58,7 let, viz Tab. 4). Z důvodu vyloučení faktoru rozdílného průměrného věku byl věk individuálně korelován na expresi souhrnné varianty $\Delta 9_{10}$ u každé ženy z obou souborů (Obr. 45).



Obrázek 45: Korelace věku žen s expresí variant s výpadky exonů $\Delta 9_{10}$ v leukocytech. Tečkovanými čarami jsou vyznačeny spojnice trendu pro každou skupinu (modře – zdravé kontroly; oranžově – pacientky s karcinomem prsu = BC pacienti)

Výsledky lineární regrese prokázaly statisticky nesignifikantní korelaci exprese ASV $\Delta 9_{10}$ u skupiny zdravých kontroly ($y = 0,0573 - 0,0001 \cdot x$; $r = -0,0918$; $r^2 = 0,0084$; $p = 0,5535$) i u skupiny pacientů s karcinomem prsu ($y = 0,0444 + 0,0002 \cdot x$; $r = 0,1953$; $r^2 = 0,0381$; $p = 0,1611$; Obr. 45). Expres variant s výpadkem exonů $\Delta 9_{10}$ je v leukocytech periferní krve signifikantně vyšší v souboru pacientek s karcinomem prsu o necelých 30 % a zároveň není závislá na věku vyšetřovaných žen.

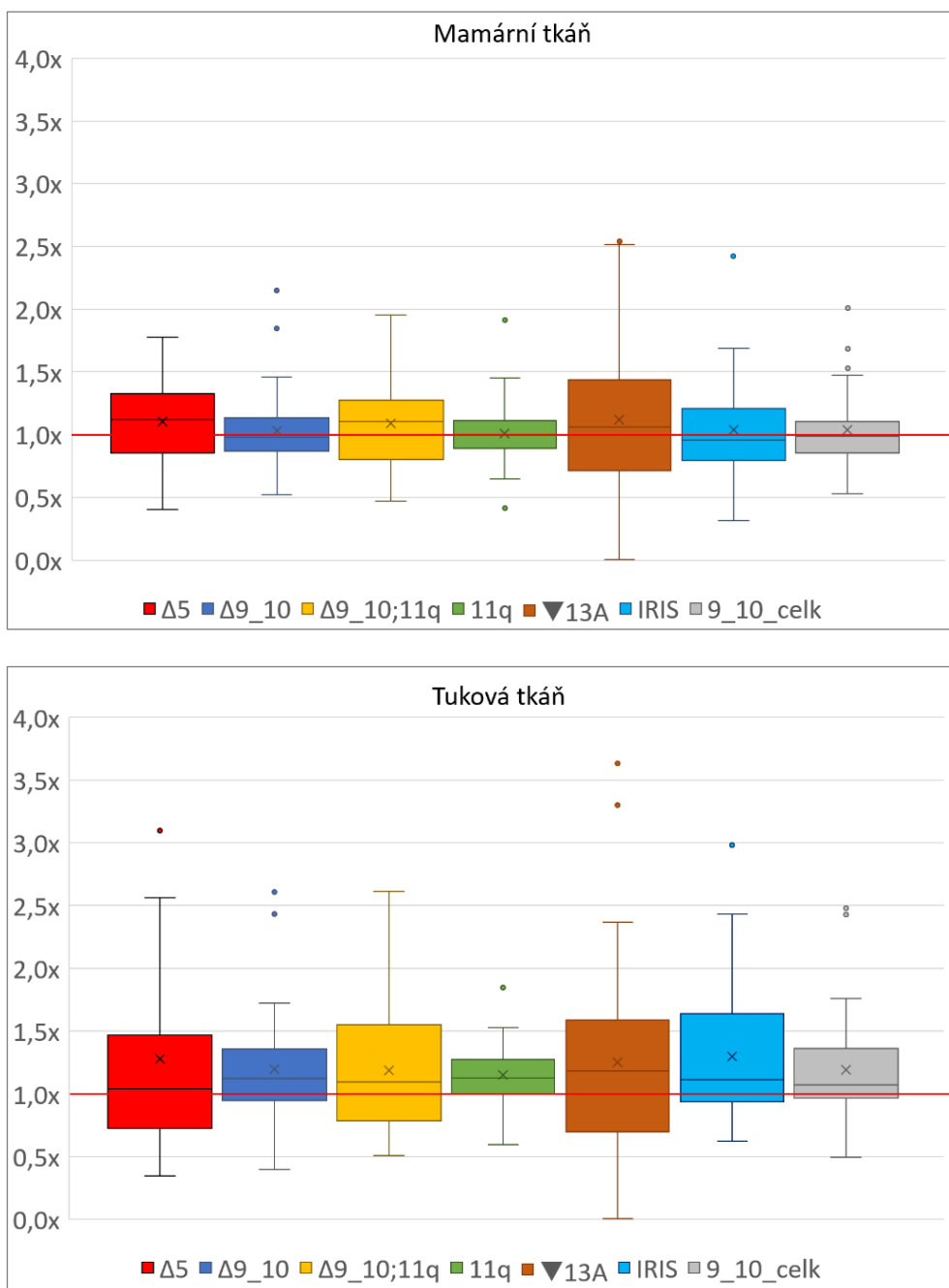
4.3.6 Laterálně specifická exprese jednotlivých variant

Od asymptomatických pacientek s mutacemi v *BRCA1*, které podstoupili profylaktické mastektomie a od zdravých kontrol, které podstoupily kosmetické operace, byl získán materiál z obou prsních (40 žen) a přilehlých tukových tkání (43 žen). Bylo provedeno porovnání laterálně specifické míry exprese jednotlivých sestřihových variant v obou vyšetřovaných tkáních i individuálně u každé ženy (Tab. 13).

Tabulka 13: Statistické porovnání laterální exprese jednotlivých variant. Poměr exprese vyjadřuje násobek exprese v levé tkáni oproti pravé tkáni. Expres jednotlivých variant je vyjádřena v % celkové exprese BRCA1 (e7). SD = směrodatná odchylka; SEM = střední chyba průměru. Pro statistické porovnání byl použit neparametrický Mann-Whitney test. Vypočítané P hodnoty jsou označeny: **červeně** $p > 0,0073$ a **zeleně** $p < 0,0073$. Hodnota $p < 0,0073$ byla použita z důvodu eliminace chyby typu I.

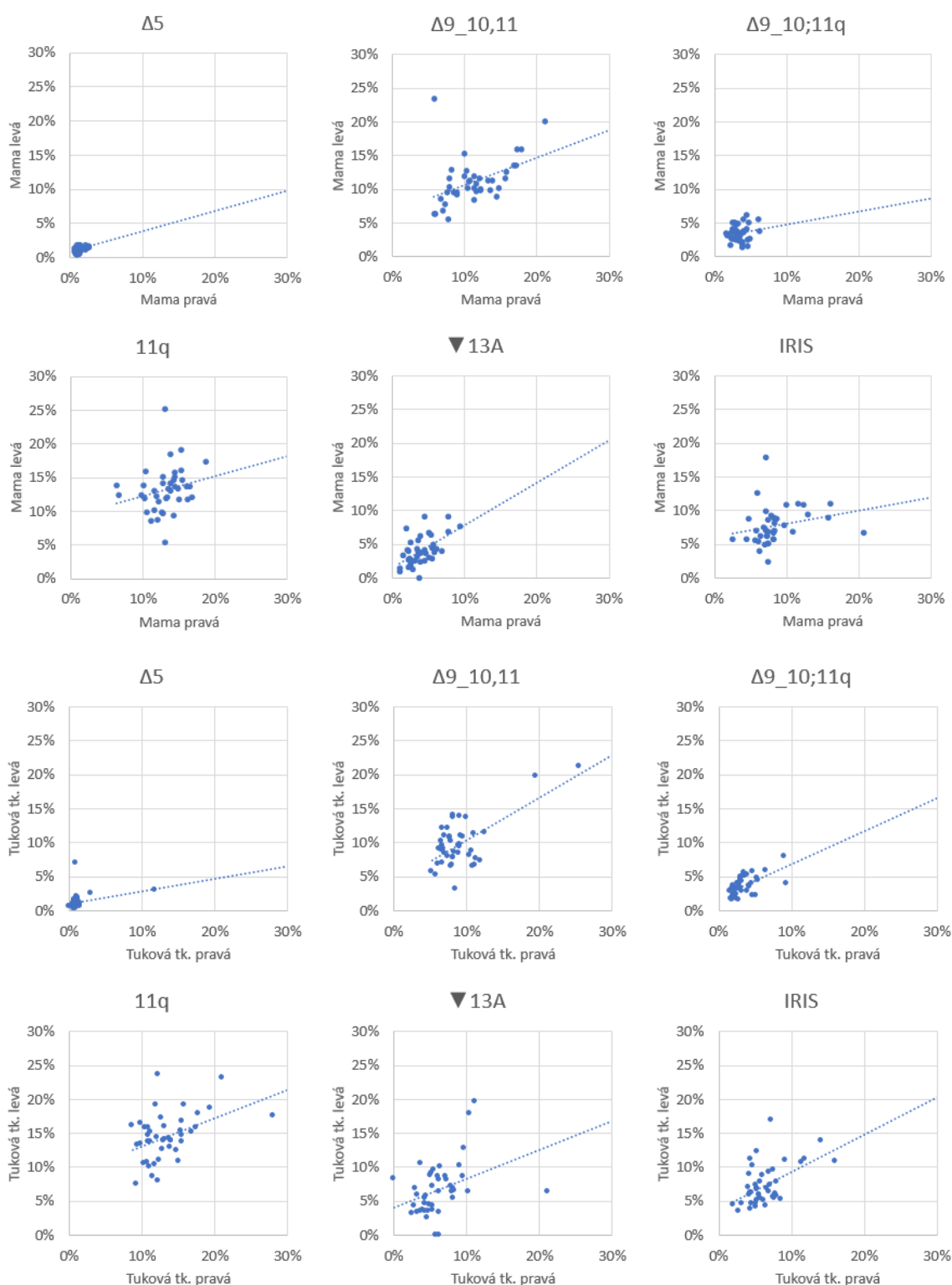
Název			$\Delta 5$	$\Delta 9_{10}; 11$	$\Delta 9_{10}; 11q$	11q	▼13A	IRIS	9_10 celková
Mamární tkáň	Pravá	Expres	1,17%	11,17%	3,39%	13,01%	3,89%	8,41%	14,55%
		SD	0,57%	3,75%	1,04%	2,49%	1,89%	3,34%	4,44%
		SEM	0,09%	0,59%	0,16%	0,39%	0,30%	0,53%	0,70%
	Levá	Expres	1,18%	11,20%	3,46%	13,20%	4,05%	7,76%	14,65%
		SD	0,40%	3,43%	1,16%	3,34%	2,13%	2,71%	4,09%
		SEM	0,06%	0,54%	0,18%	0,53%	0,34%	0,43%	0,65%
	Poměr exprese		1,123	0,977	1,107	1,003	1,061	0,956	0,986
	P		0,3240	0,9578	0,8512	0,8361	0,7327	0,3786	0,8890
Tuková tkáň	Pravá	Expres	1,26%	9,04%	3,46%	13,41%	6,24%	6,46%	12,49%
		SD	1,70%	3,53%	1,72%	3,59%	3,34%	2,78%	4,69%
		SEM	0,26%	0,54%	0,26%	0,55%	0,51%	0,42%	0,72%
	Levá	Expres	1,25%	9,74%	3,59%	14,48%	6,77%	7,40%	13,34%
		SD	1,06%	3,41%	1,41%	3,47%	3,84%	2,95%	4,41%
		SEM	0,16%	0,52%	0,21%	0,53%	0,59%	0,45%	0,67%
	Poměr exprese		1,039	1,122	1,096	1,127	1,183	1,114	1,069
	P		0,6440	0,1160	0,3248	0,0499	0,3691	0,1140	0,1385

Statistická analýza laterálně specifické exprese ASV BRCA1 v porovnání pravé a levé mamární a přilehlé tukové tkáně neodhalila odlišnou expresi, která by byla spolehlivě signifikantní. Expres jednotlivých variant je obdobná (Obr. 46).



Obrázek 46: Relativní exprese jednotlivých variant v porovnání levé vs. pravé mamární a tukové tkáně. Červenou linkou je vyznačena exprese v pravé tkáni (1x), Barevně je zobrazena exprese levé tkáně.

Pro porovnání závislosti byly vyneseny individuální hodnoty exprese z pravé a levé tkáně konkrétní varianty do grafů a poté na základě korelační analýzy porovnány (Obr. 47).



Obrázek 47: Grafy laterálních závislosti relativních expresí jednotlivých variant. Hodnoty jsou vztažené k % celkové exprese BRCA1 vyjádřenou e7. Soubory obsahovaly 40 laterálních vzorků mamární tkáně a 43 laterálních vzorků tukové tkáně. Modrými body jsou vyneseny jednotlivé hodnoty, modrá tečkovaná přímka vyjadřuje spojnici trendu grafů.

Byla prokázána statisticky signifikantní závislost exprese u všech variant s výjimkou varianty $\Delta 9_{10,11q}$; varianty 11q a varianty IRIS v mamární tkáni a varianty $\Delta 5$ s nízkou relativní expresí v tukové tkáni (Tab. 14).

Tabulka 14: Výsledky korelační analýzy laterální exprese jednotlivých variant v mamární či přilehlé tukové tkáni. Červeně jsou zvýrazněny statisticky signifikantní hodnoty ($p < 0,05$)

Název	$\Delta 5$	$\Delta 9_{10};$ 11	$\Delta 9_{10};$ 11q	11q	▼13A	IRIS	$\Delta 9_{10}$ celk.
Mama	0,424	0,443	0,173	0,221	0,561	0,239	0,417
Tuková tk.	0,294	0,648	0,595	0,431	0,370	0,519	0,724

Statistické porovnání pravé vs. levé mamární a pravé vs. levé tukové tkáně neprokázalo signifikantně odlišnou míru relativních expresí vyšetřovaných variant. Laterální exprese vyšetřovaných variant pozitivně koreluje, ve většině případů statisticky signifikantně.

5 DISKUZE

S příchodem a zavedením sekvenování nové generace (NGS) se mnohonásobně zefektivnily možnosti rutinní diagnostiky nukleových kyselin u vzorků od pacientů. V jednom sekvenačním běhu je dnes možno sekvenovat celou řadu genů / transkriptů, exomů / transkriptomů, nebo celých genomů, epigenomů apod., a na základě těchto dat odhalovat řady odchylek napříč vyšetřovanou tkání. NGS se dále vyvíjí a s rostoucí přístrojovou kapacitou úměrně roste i množství identifikovaných odchylek / variant, ať už na úrovni sekvence DNA/RNA nebo na úrovni exprese. Interpretace velkého množství nově nalezených variant je přitom časově náročná, a tak se společně s NGS rozvíjí biostatistika zpracování a charakterizace dat, bez které by nebylo možné nalezené varianty třídit a řadit dle důležitosti a dopadu na organismus. I přes rostoucí databáze nalezených variant a veškeré pokroky v biostatistických nástrojích (hlavně v predikci významu variant vyskytujících se v kódující části genomu), je dnes stále velmi obtížné spolehlivě charakterizovat většinu intronových variant (nejen) z důvodu vysoké heterogenity genomu v těchto oblastech.

Z těchto důvodů byly v předkládané práci vytyčeny cíle vytvořit robustní systém s vysokou senzitivitou a reprodukovatelností pro rychlou a komplexní identifikaci všech ASV mRNA jakéhokoliv zájmového genu a validovat ho na modelu *BRCA1*. Na základě takto vzniklého katalogu fyziologicky se vyskytujících ASV je možné, po vyšetření pacientů pomocí RNA-Seq metod, nepřímo odhalovat přítomnost (nejen) intronových mutací, které vedou ke kvalitativnímu či kvantitativnímu ovlivnění genové exprese na úrovni změny sestřihu pre-mRNA, a tím významně zjednodušit a zrychlit detekci a charakterizaci těchto mutací, anebo přímo, spolehlivě, identifikovat patogenní ASV, jak bylo ukázáno v pilotní studii konsorcia ENIGMA (Brandão et al., 2019).

Předchozí experimenty s identifikací ASV, které byly provedeny na našem pracovišti (Hojný, 2012), byly dle tehdejšího vzoru (Orban a Olah, 2001 a 2003) založeny na amplifikaci celého transkriptu *BRCA1* pomocí PCR s primery lokalizovanými do 5' a 3' UTR, případně pomocí primerů rozdělujících transkript na část před dlouhým exonem 11 a za dlouhým exonem 11. Výsledkem byla identifikace „pouze“ 13 ASV *BRCA1* a poznatky, že takto navržený metodický postup se vzorky zdravé lidské tkáně vykazuje nízkou citlivost a reprodukovatelnost z důvodů: a) obecně nízké exprese *BRCA1*; b) majoritní amplifikace referenčních variant *BRCA1* wt a *BRCA1* 11q; c) preferenční amplifikace velmi krátkých variant. Dalším negativním aspektem byly vysoké nároky na kvalitu a množství izolovaného materiálu

(integrita RNA) a časová náročnost celé metodiky (několikanásobná PCR, izolace a reamplifikace unikátních amplikonů z agarózového gelu, restriční analýzy, Sangerovo sekvenování atd.).

Předkládaná metodika založená na mPCR amplifikaci krátkých exon-exonových spojení a jejich sekvenování pomocí NGS vychází ze zkušeností předchozí práce. Navržená a zavedená multiplexní PCR, u které jsou cílovým výsledkem unifikované amplikony o velikosti okolo 80 párů bází: a) eliminuje preferenční amplifikaci krátkých ASV v reakci; b) díky vynechanému primeru v následujícím kanonickém exonu eliminuje amplifikaci exon-exonových spojení, která se vyskytují ve wt variantách BRCA1 transkriptů; c) výrazně snižuje náročnost na množství a kvalitu izolované RNA, protože pomocí hexanukleotidů se i v případě zlomy poškozeného mRNA transkriptu syntetizuje cDNA po celé jeho délce. Finální sekvenování mPCR produktů pomocí sekvenování nové generace s hlubokým pokrytím (s předchozím obohacením analyzovaného vzorku o požadované krátké amplikony pomocí velikostní selekce) zajišťuje vysokou citlivost metody.

Veškeré dílčí kroky navržené metodiky identifikace ASV se dají v dnešní době označit za dobře známé, standardně prováděné a ve většině případů rychlé. NGS je už několik let rutinně praktikováno v provozech nemocnic či soukromých genetických zařízení, postupy pro tvorbu sekvenačních knihoven (současně s velikostní selekcí) jsou rozšířené, komerčně dostupné a v mnoha případech vzájemně zaměnitelné. Jediný krok metodiky, který je oproti ostatním krokům časově náročnější a není součástí řady rutinních provozů, je navržení, optimalizace a ověření funkčnosti primerů pro mPCR reakci. Časová náročnost zde primárně závisí na struktuře analyzovaného genu. Čím více má gen exonů, tím více je nutné navrhnout primerů a provést optimalizačních a ověřujících reakcí. Samotná analýza ASV (izolace RNA → syntéza cDNA → mPCR → velikostní selekce → příprava sekvenační knihovny → sekvenování → analýza dat) je pak již relativně snadnou záležitostí.

Negativním aspektem zvolené metodiky je ztráta informace o vzájemné kombinaci nezávislých sestřihových událostí v transkriptu a ztížená komplexní identifikace intronových exonizací do transkriptu. Pomocí navržené metodiky je relativně snadné identifikovat krátké inserce přilehlé k exonům (např. varianty 1aA: c.-20+1_-20+89ins89; 10a: c.594-21_594-1ins21 či 16a: c.4986+1_4986+65ins65), nicméně exonizace vnitřních částí intronů mohou být touto metodou problematicky detekovatelné, zvláště pokud se bude jednat o exonizace dlouhých úseků DNA. Většina dlouhých mPCR amplikonů je totiž vyloučena na základě velikostní selekce, a zároveň použitý sekvenační kit se čtením 75 bází z každé strany DNA

inzertu neumožňuje u dlouhých inzertů úplnou identifikaci. Nabízejícím se vylepšením této metodiky by mohlo být prodloužení času amplifikační části mPCR a velikostní selekce ve více krocích, jejímž výsledkem bude více různě dlouhých frakcí mPCR produktů, které by se daly analyzovat sekvenačními kity s různou délkou čtení. To ovšem na úkor rychlosti a celkové ceny, která se dvěma sekvenacemi téměř zdvojnásobí. I přes toto navrhované vylepšení bychom však nezískali informaci o kombinaci sestřihových událostí na úrovni celého transkriptu, a to z důvodu omezené délky cDNA fragmentů získaných syntézou za pomoci náhodných hexanukleotidů. Analýza celých, zvláště pak dlouhých, transkriptů je vzhledem k jejich fragilitě v dnešní době stále omezena. Na úrovni cDNA je k tomu nutno přičíst nízké efektivity reverzních transkriptáz a na úrovni nativní RNA pak velmi vysoké nároky na finance z důvodu teprve se vyvíjejících technologií tzv. sekvenování třetí generace (neboli sekvenování dlouhých fragmentů).

Funkčnost a vysoká senzitivita navržené metodiky byla ověřena na modelu ASV BRCA1. Ve vyšetřovaných tkáních bylo touto metodou identifikováno 94 různých alternativních sestřihových variant / událostí BRCA1 mRNA (viz Tab. 6), což je doposud nejvíce popsanych variant. Tato práce navíc jako první hodnotí ASV BRCA1 ve fyziologických tkáních větších souborů žen. Souhrnná práce konsorcia ENIGMA z roku 2014, založená na PCR metodách v kombinaci s relativně citlivou fragmentační analýzou odhalila „pouze“ 63 ASV BRCA1 (viz Tab. 3; Colombo et al., 2014) a to většinou na imortalizovaných buněčných liniích či derivátech krevních buněk.

Ze všech variant, identifikovaných v této práci, jich dosud nebylo popsáno 42, přičemž pět z nich bylo vyhodnoceno jako predominantní – tři exonizace vnitřní části intronů ▼145bp intron 2; ▼97bp intron 8 a ▼129bp intron 21, které způsobují posun čtecího rámce a dvě varianty využívající volná sestřihová místa, které obě čtecí rámce zachovávají: 6q (SDSΔ – del 9 bp) a 10a (SAS▼ – inserce 21 bp; viz Tab. 6). Míra exprese varianty 10a se v porovnání s odvozenou wt variantou, pohybovala ze všech nově identifikovaných predominantních variant nejvýše (3,7 % – 30,5 %). Zvláště vysoké hodnoty vykazovala v leukocytech zdravých kontrol (19,5 %) a leukocytech (30,5 %) i prsní tkáni (29,2 %) pacientů s karcinomem prsu (viz Obr. 33 a příloha 3). Tato ASV zapříčiní vložení 7 dodatečných aminokyselin mezi exony 9 a 10, a ačkoliv v této oblasti není kódována žádná známá funkční doména proteinu (viz kapitola 1.3.2.1 *Struktura transkriptu BRCA1*), vysoká exprese této varianty v prsní tkáni u patientek s karcinomem prsu (29,2 % oproti 6,7 % u zdravých kontrol) je překvapivá a činí z této varianty vhodného kandidáta pro případné funkční analýzy ASV BRCA1.

V práci Colombo et al., 2014, je popsáno 11 variant, které nebyly v této práci zvolenou metodikou detekovány. Jedná se o 4 varianty s „alternativním“ nekódujícím exonem 1B, variantu BRCA1 IRIS s alternativním ukončením transkriptu, čtyři varianty multikazetové delece $\Delta 14_{-18}$; $\Delta 14_{-19}$; $\Delta 21_{-23}$ a $\Delta 22_{-23}$ a dvě varianty s alternativně zkráceným exonem 1A v kombinaci s částečnou delecí exonu 2 ($\Delta 1Aq_{-}\Delta 2p$) a delecí exonu 3 v kombinaci s inzercí alternativního exonu 4 ($\Delta 1Aq_{-}3, \blacktriangledown 4$). Varianty s alternativním exonem 1B a „speciální“ transkript BRCA1 IRIS nebyly na základě návrhu primerů pro mPCR pokryty, a proto nemohly být detekovány. Existence transkriptu s exonem 1B, který byl součástí sekvence NM_007295 (Orban a Olah, 2003), je pochybná – tato sekvence byla odstraněna z databází pro nedostatek důkazů transkripce, z tohoto důvodu nebyla tato varianta zahrnuta do návrhu mPCR primerů. O existenci zvláštní varianty BRCA1 IRIS není sporu (ElShamy a Livingston 2004), a proto byla rovnou zahrnuta do navazujících RT-qPCR analýz. Zbývajících 6 nedetekovaných variant se, vzhledem k vysoké citlivosti metody a detekci řady obdobných variant, s nejvyšší pravděpodobností ve zkoumaných fyziologických vzorcích nenacházelo.

Odlišný proteinový produkt může potenciálně tvořit 11 z 29 identifikovaných, predominantních ASV. Pět z těchto variant bylo přímo kvantifikováno na základě NGS analýz. Varianta 10a, diskutovaná výše a čtyři další ASV vznikající na základě posunu sestřihového místa, vedou k výpadku několika AK z transkriptu: 6q (poslední 3 AK v exonu 6); 8p (první AK v e8); 13p (první AK v exonu 13) a 14p (první AK v exonu 14). Oblasti variant 6q a 8p nekódují žádnou známou funkční doménu a s nejvyšší pravděpodobností tyto delece nemají vliv na funkci proteinu BRCA1. Ačkoliv varianty 13p a 14p postihují oblast domény bohatou na serinové zbytky, která hraje významnou úlohu v signalizaci (viz kapitola 1.3.2.1 *Struktura transkriptu BRCA1*), konkrétní chybějící aminokyseliny nejsou lokalizovány v místech spojených s fosforylací BRCA1 (Ouchi et al., 2006). Vzhledem k četnostem detekovaných variant, které se pohybují mezi 5,4 % až 12,4 % (13p) a 11,1 % až 26,6 % (14p) napříč všemi vyšetřovanými lidskými tkáněmi s podobnou mírou exprese u zdravých žen a pacientek s karcinomem prsu, nepředpokládáme negativní vliv u těchto variant na funkci BRCA1. Naopak, v těchto případech se může jednat o evolučně konzervovaný mechanismus, zajišťující tvorbu funkčního proteinu i přes mutaci konkrétního konsenzního sestřihového místa.

RT-qPCR analýzami bylo dále kvantifikováno 10 predominantních variant, či variant vyskytující se ve vysoké míře v 6 z 9 analyzovaných lidských tkání, které zachovávají čtecí rámec. Z těchto byly na základě hrubých Cp dat vyřazeny varianty $\Delta 3_{-}5$ a $\Delta 11$, jejichž exprese může být odhadnuta na < 0,01% exprese celkové BRCA1. Na základě úvodních NGS analýz

při identifikaci ASV BRCA1 v lidských tkáních vykazovaly obě varianty do 500 identifikovaných čtení (normalizovaná čtení na celkový 1 milión čtení; viz Tab. 6). Dále byly z RT-qPCR analýz vyřazeny varianty $\Delta 3$ a $\Delta 9_{-11}$, které se pohybovaly mezi 0,01 % a 1 % exprese z celkové BRCA1 (viz Obr. 38), z důvodu překročení spolehlivého detekčního limitu metody, který se pohybuje okolo 1 % exprese celkové BRCA1 (Hojný, 2012). Varianta $\Delta 3$ vykazovala v rámci prvotní NGS analýzy 2135 až 10293 normalizovaných čtení a varianta $\Delta 9_{-11}$ pouze množství v řádu stovek čtení, přitom jejich reálná exprese odvozená z RT-qPCR na celém vyšetřovaném souboru je obdobná (avšak s velmi vysokým rozptylem).

BRCA1 transkripty $\Delta 3$ a $\Delta 3_{-5}$ postrádají kódující sekvenci podstatné části N koncové RING domény, která je podstatná pro tvorbu E3-ubikvitinligázového heterokomplexu s proteinem BARD1, zodpovědného za ko-transport BRCA1 do jádra (Morris a Solomon, 2004) a varianty $\Delta 9_{-11}$ a $\Delta 11$ přicházejí kvůli výpadku dlouhého exonu 11 o více než polovinu sekvence referenčního transkriptu, společně s DNA vazebnou doménou, která je v exonu 11 kódována. Ačkoliv nám není známa práce, která by funkčně charakterizovala proteiny vznikající z těchto variant, můžeme se vzhledem k výraznému narušení struktury kanonického transkriptu domnívat, že vznikající izoformy mohou vykazovat výrazně odlišnou, pravděpodobně sníženou, funkční kapacitu v porovnání s popisovanými funkcemi wt BRCA1. Na základě této domněnky by pak nebylo překvapivé, že se tyto varianty ve zdravých tkáních vyskytují ve velmi nízkém procentu (< 1%).

Zbylé varianty zahrnuté do RT-qPCR analýzy $\Delta 5$; $\Delta 9_{-10}$; $\Delta 9_{-10,11q}$; $11q$; $\nabla 13A$ a IRIS vykazovaly relativní expresi vyšší než 1 % z celkového množství BRCA1 transkriptu a byly plně zhodnoceny.

Z plně hodnocených variant byla nejméně exprimována ASV $\Delta 5$, která postrádá podstatnou část kódující sekvence RING domény. Vazebná RING doména s E3 ubikvitinylací aktivitou zajišťuje vazbu důležitého vazebného partnera BRCA1 – proteinu BARD1, který je zásadní pro enzymatickou aktivitu komplexu, a zároveň pro lokalizaci proteinu BRCA1 do jádra (vazba s BARD1 umocňuje lokalizaci a kryje jaderné exportní signály BRCA1; Rodríguez a Henderson, 2000; Fabbro et al., 2002). Vzhledem k předpokládanému funkčnímu omezení vzniklé izoformy je tedy překvapivé, že míra exprese této varianty odpovídá 1,10 % – 1,13 % celkové exprese BRCA1 ve všech vyšetřovaných tkáních a neliší se signifikantně u souboru zdravých žen a žen s karcinomem prsu (NGS analýza této varianty identifikovala 4013 až 13043 normalizovaných čtení). Relativní exprese transkriptu BRCA1 $\Delta 5$ je uniformní a nízká napříč všemi vyšetřovanými tkáněmi i skupinami, vykazuje velmi malý rozptyl naměřených hodnot a

potencionálně vznikající proteinový produkt s nejvyšší pravděpodobností neplní funkci wt formy BRCA1, protože nemůže být lokalizován do jádra. Z těchto důvodů bychom se mohli domnívat, že varianta $\Delta 5$ může být vedlejším produktem sestřihu BRCA1 na základě určitého, opakujícího se, vzorce sestřihového aparátu, který vyplývá z jeho značné robustnosti. Se svou mírou exprese 1 % z celkové BRCA1 pravděpodobně nemá vliv na výslednou funkci BRCA1 složenou ze všech identifikovaných transkriptů, či plní jinou, dosud neobjasněnou funkci v organismu.

Další variantou je známá BRCA1 11q (někdy také označována jako BRCA1a), která využívá alternativní sestřihové místo vzdálené 3309 bp od 3' konce exonu 11. BRCA1 11q byla jednou z prvních identifikovaných ASV BRCA1 (Lu et al., 1996), protože se vyskytuje v relativně vysokém množství a zkrácený transkript o více než polovinu se při amplifikaci od 5' do 3' UTR obvykle lépe amplifikuje z důvodu nižších nároků na kvalitu izolované RNA (Hojný, 2012). Ačkoliv je v postrádané sekvenci exonu 11 lokalizována DNA vazebná doména, oba jaderné lokalizační signály a část domény bohaté na serinové zbytky (viz kapitola 1.3.2.1 *Struktura transkriptu BRCA1*), je tato varianta napříč vyšetřovanými tkáněmi exprimována ve vysoké míře 11,17 % (leukocyty), 12,95 % (mamární tkáň) a 13,91 % (tuková tkáň) z celkové BRCA1. Relativní míra exprese není signifikantně odlišná u skupiny zdravých žen a skupiny žen s karcinomem prsu. NGS analýza překvapivě identifikovala výrazně rozdílné množství čtení mezi 823 a 7879 napříč vyšetřovanými tkáněmi. Funkcí transkripční varianty BRCA1 11q, respektive výsledného proteinu, se zabývala řada studií, převážně z laboratoře V.N. Rao, které souhrnně naznačují mírně odlišné funkce této sestřihové varianty od wt proteinu BRCA1. Wang a kolektiv (1997) naznačují odlišnou lokalizaci 11q v cytoplazmě spojenou s vazbou E2F cyklinů a CDKs, která by mohla způsobovat důležitou inhibici buněčného cyklu a tím zabráňovat rozvoji tumorogeneze. Práce Chaie a kolektivu (1999 a 2001) popisují schopnost přímé vazby variant s exonem 11q k proteinům p53 a Elk-1 a díky tomu schopnost inhibice růstu buněčných linií odvozených z buněk karcinomu prsu. Následující práce (Maniccia et al., 2009) dokumentuje mitochondriální lokalizaci variant s exonem 11q a další antiproliferativní účinky na buněčné linie karcinomu prsu. Série dalších prací popisují odlišnou vazebnou kapacitu BRCA1 11q k E2 konjugáčnmu enzymu Ubc9, která může mít dopad na aktivitu estrogenního receptoru alfa (Xu et al., 2009), regulaci růstu buněčných linií karcinomu ovaria (Qin et al., 2012; Xu et al., 2016) a lokalizaci Calveolinu 1 (Xu et al., 2014).

Většina prací, zmíněných v předchozím odstavci, zároveň zkoumá se stejnými nebo velmi podobnými výsledky i „odvozenou“ variantu $\Delta 9_10,11q$, která navíc postrádá exony 9 a 10 a

kteřá je v literatuře často označována jako BRCA1b (Lu et al., 1996). Tato sestřihová varianta se vyskytuje v mamární a přilehlé tukové tkáni v signifikantně vyšším poměru celkové exprese BRCA1 nežli v leukocytech periferní krve (3,27 % a 3,21 % vs 1,22 %; $p = 0$). Na rozdíl od mamární a přilehlé tukové tkáně vykazuje tato varianta v leukocytech periferní krve vyšší expresi u pacientů s karcinomem prsu oproti zdravým ženám o 25,9 % (s hraniční signifikancí $p = 0,048$). NGS analýza pro tuto konkrétní variantu není k dispozici, odráží pouze souhrnnou variantu $\Delta 9_10$.

ASV $\Delta 9_10$ s dlouhou formou exonu 11 se ve vyšetřovaných tkáních exprimuje ve vyšší míře než „příbuzná“ $\Delta 9_10,11q$, nicméně opakuje podobný vzorec tkáňově specifické exprese (mamární tkáň 10,21 %; přilehlá tuková tkáň 9,11 % a leukocyty 3,96 %; $p = 0$ v obou případech). Navíc tato varianta vykazuje u leukocytů periferní krve vyšší relativní míru exprese o 29,8 % u pacientek s karcinomem prsu oproti zdravým kontrolám ($p < 0,0073$), zatímco v mamární tkáni je exprese u pacientek s karcinomem prsu o 14,9% nižší, než u zdravých kontrol ($p < 0,001$). Tyto výsledky potvrzuje i statistické porovnání souhrnu obou variant s výpadkem exonů 9 a 10 (pacientky mají v leukocytech periferní krve vyšší expresi této varianty o 28,8 % než kontroly; $p < 0,0073$; a v prsní tkáni o 11,7 % nižší; $p < 0,0073$). NGS analýza pro souhrnné varianty $\Delta 9_10$ vykazovala opět významný rozptyl 562 až 5109 normalizovaných čtení v závislosti na vyšetřované tkáni.

Nejčastěji detekovanou a popsanou exonizací intronové oblasti je varianta s inzercí popisovaného alternativního exonu ▼13A (Fortin et al., 2005). Tato „*in-frame*“ varianta vykazovala významně odlišnou míru exprese v prsní tkáni (4,14 % z celkové BRCA1 oproti 6,7 % v leukocytech a 6,44 % v tukové tkáni; $p = 0$) a relativní exprese se nelišila u pacientek s karcinomem prsu oproti zdravým kontrolám. NGS analýza identifikovala 703 až 7267 normalizovaných čtení napříč vyšetřovanými tkáněmi. Výsledkem inserce 66 bází dlouhého alternativního exonu 13 do transkriptu je přidání 22 AK do středu oblasti bohaté na serinové zbytky, která na základě fosforylací od signálních kináz pravděpodobně moduluje vazebnou kapacitu BRCT domén. Vliv takovéto inserce na funkčnost proteinu BRCA1 je nejasný, funkční analýza proteinu této ASV by mohla objasnit vliv na funkci wt BRCA1.

Vzhledem k diskutované literatuře a na základě relativně vysoké míry fyziologického výskytu sestřihových variant s krátkou formou exonu 11, s výpadky exonů 9 a 10 či se zahrnutým alternativním exonem 13, se nedomníváme, že by izoformy vznikající z těchto variant měly významně negativní vliv na funkci BRCA1 jako tumor supresoru. Kromě vysoce konzervovaných C a N koncových domén nemá BRCA1 ve vnitřní části proteinu významně

uspořádanou oblast. Vnitřní část BRCA1 pravděpodobně slouží hlavně jako spojník koncových domén, jak je tomu u celé řady proteinů zapojených v signalizačních procesech. Z tohoto pohledu by se dalo uvažovat jen o velmi málo omezené funkčnosti izoform, vznikajících ze zmiňovaných variant, protože všechny tyto izoformy konzervované RING, SCD a obě BRCT domény zachovávají. Struktura proteinů vznikajících z těchto variant se liší ve střední části délkou primárního řetězce, a vzhledem k tomu, že dosud není známa 3D struktura střední části BRCA1, se nedá vliv výpadků či inzercí jmenovaných ASV detailněji posoudit.

Důležitý význam variant 11q a $\Delta 9_10$ se pravděpodobně projevuje až v případě výskytu „stop“ či „frameshift“ mutací lokalizovaných v úsecích exonů 9, 10 a dlouhé 11, které sestřihový aparát s využitím alternativního sestřihu, eliminuje. Seo a kolektiv (2018) prokázali ve dvou případech rodin, nesoucích mutace vedoucí k předčasnému STOP kodonu v dlouhé části exonu 11, zvýšený výskyt varianty 11q (společně se snížením exprese či nepřítomnosti transkriptu o plné délce obsahující terminační kodon, pravděpodobně na základě mechanismu nonsense-mediated decay), která byla do určité míry schopna nahradit funkci wt BRCA1, resp. výrazně omezit negativní efekt nepřítomnosti BRCA. Tento mechanismus se pravděpodobně může uplatnit i v případě variant s truncačními mutacemi v exonech 9 a 10.

Poslední kvantifikovanou variantou je neobvyklá varianta BRCA1 IRIS, která na konec svého transkriptu inkorporuje přilehlou část intronu 11 s vlastním poly-A signálem, což vede k „předčasnému“ ukončení translace. Kvantifikace této varianty prokázala výrazně zvýšenou relativní míru exprese v leukocytech (14,62 %) oproti mamární (7,16 %, $p = 0$) a přilehlé tukové tkáni (6,24 %; $p = 0$). Dále byla prokázána signifikantně nižší míra exprese této varianty u pacientek s karcinomem prsu oproti zdravým kontrolám o 15,4 % ($p < 0,0073$). BRCA1 IRIS byla popsána v roce 2004 (ElShamy a Livingston 2004), a od této doby je v laboratoři autorů intenzivně zkoumána. Autoři v řadě svých prací postupně prokazují pravděpodobně protumorigenní vlastnosti této varianty v případě její overexprese, která byla pozorována u karcinomů prsu a plic a také u akutní myeloidní leukémie. BRCA1 IRIS má schopnost vazby na chromatin a ovlivnění transkripce řady genů. V případě vštěpení patientských nádorových buněk s overexprimovanou IRIS do myších modelů je pozorována schopnost metastazování těchto buněk na základě ovlivnění exprese na úrovni transkripce (inhibice PTEN vedoucí k aktivaci HIF1 α ; Li et al., 2017). Určení fyziologické míry exprese této ASV je tedy důležitým předpokladem pro správné posouzení overexprese, což by mohl být jeden z markerů určitých typů nádorových onemocnění.

Výsledky kvantifikací provedených v této práci potvrdily obecně nízkou expresi *BRCA1*. Ačkoliv experiment nebyl navržen pro absolutní kvantifikaci celkové *BRCA* mRNA, z porovnání s referenčními geny *GAPDH*, *UBC* a *ATP5B* vyplývá 100x až 1000x nižší exprese *BRCA1* (viz Obr. 36), což vzhledem k tumor-supresivní úloze proteinového produktu není, v případě zdravých tkání dospělých jedinců překvapivé, a je ve shodě s RNA-seq daty databáze GTEx (<https://www.gtexportal.org>).

Z porovnání výsledků relativních expresí oblastí zastupujících veškerou *BRCA1* (v exonu 7 a v oblasti exonů 23-24) s referenčními geny, vyplývá vyšší relativní exprese amplikonu lokalizovaném v e7 (viz Obr. 37). Tato zvýšená exprese naznačuje přítomnost transkriptů, postrádajících exony 23 a 24, tedy pravděpodobně transkriptů s alternativně ukončenou translací, jako je například *BRCA1* IRIS. Samotná exprese ASV IRIS tyto značné rozdíly, zvláště v leukocytech periferní krve (až dvojnásobné), nevysvětluje, protože je v leukocytech exprimována jen o cca 7–8 % více v porovnání s ostatními tkáněmi. Tento výsledek naznačuje přítomnost další varianty či variant postrádajících 3' koncovou část transkriptu. Tyto hypotetické varianty jsme nebyli schopni zvoleným NGS přístupem, zaměřeným na analýzu exon-exonových spojení, detekovat.

NGS přístup identifikace ASV variant umožnil omezeným způsobem kvantifikovat nalezené ASV varianty na základě množství identifikovaných čtení normalizovaných na celkový počet čtení ve vzorku. Při srovnání NGS „kvantifikace“ s přesnou RT-qPCR metodou vyšla najevo některá omezení NGS „kvantifikace“. Např. u ASV $\Delta 3$ a $\Delta 5$, s mírou exprese okolo 0,1 % a 1,1 % celkové *BRCA1*, bylo NGS přístupem identifikováno ve zdravých tkáních leukocytů 3120, v prsní tkáni 4141 a v tukové tkáni 2135 čtení ($\Delta 3$), resp. 8016, 5524, 4778 ($\Delta 5$). Přičemž u relativně hodně exprimovaných variant $\Delta 9_10$ a 11q s expresí až 10 a více % celkové *BRCA1*, bylo identifikováno v leukocytech 562, v prsní tkáni 998 a v tukové tkáni 940 čtení ($\Delta 9_10$), resp. 823, 2971 a 1048 čtení u 11q. Nízká korelace NGS „kvantifikace“ je pravděpodobně způsobena vstupem pouze 8 vzorků do fáze identifikace (mPCR), které byly navíc normalizovány na základě množství celkové RNA, nikoliv na základě množství *BRCA1* mRNA vstupující do reakce, a dále na základě preferenční amplifikace, která se při 35 cyklech + 6 cyklech PCR při přípravě knihoven pravděpodobně projevila, a to i přes snahu navrhnout amplikony mPCR tak, aby byla jejich velikost v případě alternativně sestřižených transkriptů co nejvíce podobná. Na druhou stranu ostatní varianty, které byly z kvantifikace vyřazeny na základě C_p hodnot za spolehlivým detekčním limitem qPCR metody ($\Delta 3_5$, $\Delta 11$, $\Delta 9_11$) správně vykazovaly mnohem nižší hodnoty množství čtení identifikovaných NGS (> 500 čtení).

Omezená kvantifikace ASV se dá na základě NGS analýz provést, ale pro přesné výsledky musí být použity metody kvantitativní PCR či ddPCR.

I přes zmiňované nedostatky v kvantifikaci byla navržená metodika identifikace ASV genu našeho zájmu v našich laboratořích již několikrát úspěšně využita a průběžná, či pre-finální, data těchto experimentů vykazují výborné výsledky.

6 ZÁVĚR

Alternativní sestřih a varianty vznikající na jeho základě jsou více než 30 let intenzivně zkoumány. Během této doby byla objasněna řada mechanismů, které alternativní sestřih způsobují nebo jinak ovlivňují. Současně byla na řadě konkrétních případů demonstrována důležitá a nezastupitelná role ASV v organizmu. Nemálo ASV bylo popsáno v průběhu výzkumu genů predisponujících ke karcinomu prsu, zvláště u *BRCA1*, který byl v souvislosti s tímto onemocněním identifikován a intenzivně zkoumán jako první. První ASV *BRCA1* byly popsány už v roce 1996 a od té doby jich byly identifikovány desítky. Nedlouho po objevení ASV se začaly objevovat práce zkoumající funkci identifikovaných ASV *BRCA1*, přičemž byl prokázán nesporný patogenní efekt několika z nich. Přesto nebyla otázka složení a množství jednotlivých variant v lidských tkáních dosud uspokojivě zodpovězena. Z důvodu nedostatečných znalostí o principu fungování regulačních míst alternativního sestřihu také nebylo dlouhou dobu zcela zřejmé, jaké varianty se vyskytují přirozeně a jaké se vyskytují na základě mutací, postihujících tato místa. Taktéž neexistoval levný, jednoduchý a rychlý přístup k analýze a identifikaci ASV, který by měl dostatečnou citlivost. Proto se předkládaná práce pokusila tyto problémy vyřešit.

V rámci řešení předkládané práce byl navržen přístup, který umožňuje s využitím sekvenování nové generace identifikovat na úrovni mRNA přítomné sestřihové varianty libovolného genu našeho zájmu. Zavedená metodika byla validována na modelovém genu *BRCA1*, pravděpodobně jediného takto komplexního genu, u kterého existuje katalog ASV z roku 2014. Porovnání s tímto katalogem prokázalo, že v práci předkládaný NGS přístup identifikace je citlivější, dle použité metodiky jednoznačně rychlejší a bezesporu levnější. V návaznosti na tato zjištění je aktuálně používán s výbornými průběžnými výsledky pro analýzu dalších ASV jiných genů (dosud nepublikováno).

Na základě NGS identifikace ASV *BRCA1* bylo přesnou RT-qPCR kvantifikováno několik nejčastěji se vyskytujících „*in-frame*“ variant, poprvé na větším souboru zdravých tkání leukocytů periferní krve, prsní a přilehlé tukové tkáni u zdravých žen a žen, u kterých se v průběhu života vyvinul karcinom prsu. Výsledky prokázaly, že: a) exprese *BRCA1* ve zdravých tkáních je velmi nízká; b) z celkem deseti vyšetřovaných sestřihových variant nepostihující čtecí rámec, které byly detailněji kvantifikovány metodou RT-qPCR, jich šest bylo exprimováno ve > 1% celkové *BRCA1*; c) jejich relativní exprese se pohybovala kumulativně mezi 35 – 40 % všech transkriptů *BRCA1*, bez ohledu na vyšetřovaný typ tkáně;

d) všech šest variant bylo přítomno ve všech třech vyšetřovaných tkáních, ale míra exprese některých z nich se lišila (u leukocytů periferní krve vykazuje $\Delta 9_{-10}$ více než 2x nižší a IRIS naopak 2x vyšší v porovnání s ostatními tkáněmi); e) exprese variant v prsní a přilehlé tukové tkáni spolu obvykle pozitivně, statisticky signifikantně koreluje, zatímco v porovnání s leukocyty ne; f) rozdíly v expresi vyšetřovaných ASV u souboru zdravých žen a patientek s karcinomem prsu jsou v relevantní prsní tkáni malé, statisticky signifikantně je o 15,4 % snížena exprese varianty IRIS, která má pravděpodobně odlišnou funkci než wt BRCA1 či má funkci regulační, o 11,7 % je snížena exprese $\Delta 9_{-10}$, která má pravděpodobně velmi podobnou funkci jako wt BRCA1; g) exprese variant spolu laterálně pozitivně koreluje.

Výsledky předkládaných experimentů objasňují, vzhledem k citlivosti použitých metod, složení a množství alternativních sestřihových variant BRCA1 v relevantních tkáních lidského těla – leukocytů periferní krve (ze kterých je materiál pro genetické vyšetření nejčastěji extrahován) a mamární (přilehlé tukové) tkáně, která je obvykle postižena případným karcinomem (nebo vyšetřována na aktuální somatický status *BRCA1*). Na základě kvalitativních i kvantitativních výsledků této práce, v porovnání s budoucími výsledky např. rutinních RNA-Seq analýz nádorové prsní tkáně, bude ihned možné identifikovat sestřihovou variantu vzniklou na základě aberantního sestřihu, tzn. nepřímo odhalit mutace vedoucí k aberantnímu sestřihu, anebo, na základě kvantitativní dysbalance fyziologicky se vyskytujících variant, identifikovat mutaci v regulačním sestřihovém místě.

7 SEZNAM POUŽITÝCH ZKRATEK

ASV	Alternativní sestřihová varianta
ATP5B	Beta podjednotka F1 ATP syntázy
BBP	Branch-point binding protein / vazebný protein na místo větvení
bp	Base pairs / páry bází
BRCT	BRCA1 C-terminal / BRCA1 C-koncová doména
cDNA	Komplementární DNA
Cp	Crossing point
CTD	C terminální doména
DBD	DNA-binding domain / DNA vazebná doména
ESE	Exon splicing enhancer
ESS	Exon splicing silencer
GAPDH	Glyceraldehyd-3-fosfát dehydrogenáza
hnRNP	Heterogenní nukleární ribonukleoproteiny
HRM	High resolution melting / analýza křivek tání
IGV	Integrative Genomics Viewer
ISE	Intron splicing enhancer
ISS	Intron splicing silencer
LSS	Leaky splice site / volné sestřihové místo
mPCR	Multiplexní PCR
mRNA	Messenger / mediátorová /informační RNA
NGS	New Generation Sequencing / sekvenování nové generace
PEG	Polyethylenglykol
pre-mRNA	Prekurzorová mRNA
RG	Referenční gen
RING	Really Interesting New Gene
RNAP II	DNA-dependentní RNA polymeráza II
RNA-Seq	RNA next generation sequencing / RNA sekvenování nové generace
RPKM	Reads Per Kilobase Milion / čtení vztažené k jednomu milionu čtení jedné kb transkriptu
SAS	Splice acceptor site / „příjemce“ sestřihového místa
SCD	Serine-containing domain / doména bohatá na serinové zbytky
SDS	Splice donor site / „dárce“ sestřihového místa
snRNA	Small nuclear RNA / malá jaderná RNA
snRNP	Small ribonucleoproteins particles / malé ribonukleoproteiny
SPRI	Solid Phase Reversible Immobilisation
SR	Serine-arginine rich factors / faktory bohaté na serin a arginin
SSIII	SuperScript III reverzní transkriptáza
T _{an}	Anelační teplota primerů
U2AF	U2 auxiliary factor / pomocný faktor U2
UBC	Ubikvitin C
UTR	Untranslated region / nepřekládaná oblast
VUS	Variant of uncertain signification / varianta s nejasným významem
wt	Wild type / divoký typ

8 SEZNAM POUŽITÉ LITERATURY

- Aebi M, Weissmann C. Precision and orderliness in splicing. *Trends Genet.* 1987 Apr;3(4):102-107.
- Anantha RW, Simhadri S, Foo TK, Miao S, Liu J, Shen Z, Ganesan S, Xia B. Functional and mutational landscapes of BRCA1 for homology-directed repair and therapy resistance. *Elife.* 2017 Apr 11;6. pii: e21350.
- Batsché E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol.* 2006 Jan;13(1):22-9.
- Beyer AL, Osheim YN. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 1988 Jun;2(6):754-65.
- Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.* 1997;11(1):68-76.
- Brandão RD, Mensaert K, López-Perolio I, Tserpelis D, Xenakis M, Lattimore V, Walker LC, Kvist A, Vega A, Gutiérrez-Enríquez S, Díez O; KConFaB Investigators, de la Hoya M, Spurdle AB, De Meyer T, Blok MJ. Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *Int J Cancer.* 2019 Jan 8.
- Brozova M, Kleibl Z, Netikova I, Sevcik J, Scholzova E, Brezinova J, Chaloupkova A, Vesely P, Dundr P, Zadinova M, Krasna L, Matouskova E. Establishment, growth and in vivo differentiation of a new clonal human cell line, EM-G3, derived from breast cancer progenitors. *Breast Cancer Res Treat.* 2007 Jun;103(2):247-57.
- Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* 1999 Dec;9(12):M57-60. Review.
- Cailleau R, Young R, Olivé M, Reeves WJ Jr. Breast tumor cell lines from pleural effusions. *J Natl Cancer Inst.* 1974 Sep;53(3):661-74.
- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, Livingston DM. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell.* 2001;105(1):149-160.

- Chai Y, Chipitsyna G, Cui J, Liao B, Liu S, Aysola K, Yezdani M, Reddy ES, Rao VN. c-Fos oncogene regulator Elk-1 interacts with BRCA1 splice variants BRCA1a/1b and enhances BRCA1a/1b-mediated growth suppression in breast cancer cells. *Oncogene*. 2001 Mar 15;20(11):1357-67.
- Chai YL, Cui J, Shao N, Shyam E, Reddy P, Rao VN. The second BRCT domain of BRCA1 proteins interacts with p53 and stimulates transcription from the p21WAF1/CIP1 promoter. *Oncogene*. 1999 Jan 7;18(1):263-8.
- Colombo M, Blok MJ, Whiley P, Santamariña M, Gutiérrez-Enríquez S, Romero A, Garre P, Becker A, Smith LD, De Vecchi G, Brandão RD, Tserpelis D, Brown M, Blanco A, Bonache S, Menéndez M, Houdayer C, Foglia C, Fackenthal JD, Baralle D, Wappenschmidt B; kConFaB Investigators, Díaz-Rubio E, Caldés T, Walker L, Díez O, Vega A, Spurdle AB, Radice P, De La Hoya M. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet*. 2014 Jul 15;23(14):3666-80.
- Cooper TA, Mattox W. The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet*. 1997 Aug;61(2):259-66. Review.
- Cramer P, Cáceres JF, Cazalla D, Kadener S, Muro AF, Baralle FE, Kornblihtt AR. Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*. 1999 Aug;4(2):251-8.
- Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci USA*. 1997 Oct 14;94(21):11456-60.
- Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*. 2007 Jun 22;26(6):867-81.
- de la Mata M, Kornblihtt AR. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol*. 2006 Nov;13(11):973-80.
- Dever SM, White ER, Hartman MC, Valerie K. BRCA1-directed, enhanced and aberrant homologous recombination: mechanism and potential treatment strategies. *Cell Cycle*. 2012;11(4):687-94.

- ElShamy WM, Livingston DM. Identification of BRCA1-IRIS, a BRCA1 locus product. *Nat Cell Biol.* 2004;6(10):954-967.
- Eperon LP, Graham IR, Griffiths AD and Eperon IC (1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell.* 1988 Jul;54(3), 393–401.
- Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet.* 2014 Nov 15;23(22):5866-78.
- Fabbro M, Rodriguez JA, Baer R, Henderson BR. BARD1 induces BRCA1 intranuclear foci formation by increasing RING-dependent BRCA1 nuclear import and inhibiting BRCA1 nuclear export. *J Biol Chem.* 2002;277(24):21315-24.
- Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpour S, Danielsson A, Edlund K, Asplund A, Sjöstedt E, Lundberg E, Szigartyo CA, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson A, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlén M. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* 2014 Feb;13(2):397-406.
- Filippini SE, Vega A. Breast cancer genes: beyond BRCA1 and BRCA2. *Front Biosci (Landmark Ed).* 2013 Jun 1;18:1358-72. Review.
- Fortin J, Moisan AM, Dumont M, Leblanc G, Labrie Y, Durocher F, Bessette P, Bridge P, Chiquette J, Laframboise R, Lépine J, Lespérance B, Pichette R, Plante M, Provencher L, Voyer P, Simard J. A new alternative splice variant of BRCA1 containing an additional in-frame exon. *Biochim Biophys Acta.* 2005;1731(1):57-65.
- Fu XD. The superfamily of arginine/serine-rich splicing factors. *RNA.* 1995;1(7):663-80.
- Hakem R, de la Pompa JL, Sirard C, Mo R, Woo M, Hakem A, Wakeham A, Potter J, Reitmair A, Billia F, Firpo E, Hui CC, Roberts J, Rossant J, Mak TW. The tumor suppressor gene Brcal is required for embryonic cellular proliferation in the mouse. *Cell.* 1996;85(7):1009-1023.
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990;250(4988):1684-1689.

- Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000 Jan 7;100(1):57-70. Review.
- Hojny J, Zemankova P, Lhota F, Sevcik J, Stranecky V, Hartmannova H, Hodanova K, Mestak O, Pavlista D, Janatova M, Soukupova J, Vocka M, Kleibl Z, Kleiblova P. Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model. *Gene*. 2017 Dec 30;637:41-49.
- Hojný J. Analýza a charakterizace sestřihových variant BRCA1. Praha, 2012. Diplomová práce (Mgr.). Univerzita Karlova, Přírodovědecká fakulta, Katedra buněčné biologie, 2012-06-15. <https://is.cuni.cz/webapps/zzp/download/120081082/?lang=cs>
- Horowitz DS. The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip Rev RNA*. 2012;3(3):331-50.
- Huen MS, Sy SM, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. *Nat Rev Mol Cell Biol*. 2010;11(2):138-48
- Jelen N, Ule J, Zivin M, Darnell RB. Evolution of Nova-dependent splicing regulation in the brain. *PLoS Genet*. 2007 Oct;3(10):1838-47. Epub 2007 Aug 27.
- Kadener S, Cramer P, Nogués G, Cazalla D, de la Mata M, Fededa JP, Werbach SE, Srebrow A, Kornblihtt AR. Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J*. 2001 Oct 15;20(20):5759-68.
- Kafasla P, Mickleburgh I, Llorian M, Coelho M, Gooding C, Cherny D, Joshi A, Kotik-Kogan O, Curry S, Eperon IC, Jackson RJ, Smith CW. Defining the roles and interactions of PTB. *Biochem Soc Trans*. 2012 Aug;40(4):815-20. Review.
- Kleiblova, P., Dostalova, I., Bartlova, M., Lacinova, Z., Ticha, I., Krejci, V., Springer, D., Kleibl, Z., Haluzik, M., 2010. Expression of adipokines and estrogen receptors in adipose tissue and placenta of patients with gestational diabetes mellitus. *Mol. Cell. Endocrinol.* 314, 150–156.
- Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA*. 1971 Apr;68(4):820-3.
- Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet*. 2009 Mar;41(3):376-81.
- Koonin EV, Altschul SF, Bork P. BRCA1 protein products: Functional motifs. *Nat Genet*. 1996;13(3):266-268.

- Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol.* 2013 Mar;14(3):153-65. Review. Erratum in: *Nat Rev Mol Cell Biol.* 2013 Mar;14(3).
- Kornblihtt AR. Promoter usage and alternative splicing. *Curr Opin Cell Biol.* 2005 Jun;17(3):262-8. Review.
- Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, Jervis S, van Leeuwen FE, Milne RL, Andrieu N, Goldgar DE, Terry MB, Rookus MA, Easton DF, Antoniou AC; BRCA1 and BRCA2 Cohort Consortium, McGuffog L, Evans DG, Barrowdale D, Frost D, Adlard J, Ong KR, Izatt L, Tischkowitz M, Eeles R, Davidson R, Hodgson S, Ellis S, Nogues C, Lasset C, Stoppa-Lyonnet D, Fricker JP, Faivre L, Berthet P, Hoening MJ, van der Kolk LE, Kets CM, Adank MA, John EM, Chung WK, Andrulis IL, Southey M, Daly MB, Buys SS, Osorio A, Engel C, Kast K, Schmutzler RK, Caldes T, Jakubowska A, Simard J, Friedlander ML, McLachlan SA, Machackova E, Foretova L, Tan YY, Singer CF, Olah E, Gerdes AM, Arver B, Olsson H. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA.* 2017 Jun 20;317(23):2402-2416.
- Lee JA, Tang ZZ, Black DL. An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev.* 2009 Oct 1;23(19):2284-93.
- Lhota F, Zemankova P, Kleiblova P, Soukupova J, Vocka M, Stranecky V, Janatova M, Hartmannova H, Hodanova K, Kmoch S, Kleibl Z. Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet.* 2016 Oct;90(4):324-33.
- Li AG, Murphy EC, Culhane AC, Powell E, Wang H, Bronson RT, Von T, Giobbie-Hurder A, Gelman RS, Briggs KJ, Piwnica-Worms H, Zhao JJ, Kung AL, Kaelin WG Jr, Livingston DM. BRCA1-IRIS promotes human tumor progression through PTEN blockade and HIF-1 α activation. *Proc Natl Acad Sci USA.* 2018 Oct 9;115(41):E9600-E9609.
- Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S, Weissman AM. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci USA.* 1999;96(20):11364-11369.

- Lu M, Conzen SD, Cole CN, Arrick BA. Characterization of functional messenger RNA splice variants of BRCA1 expressed in nonmalignant and tumor-derived breast cells. *Cancer Res.* 1996;56(20):4578-81.
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. Epigenetics in alternative pre-mRNA splicing. *Cell.* 2011 Jan 7;144(1):16-26. Review.
- Mallery DL, Vandenberg CJ, Hiom K. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J.* 2002;21(24):6755-6762.
- Maniccia AW, Lewis C, Begum N, Xu J, Cui J, Chipitsyna G, Aysola K, Reddy V, Bhat G, Fujimura Y, Henderson B, Reddy ES, Rao VN. Mitochondrial localization, ELK-1 transcriptional regulation and growth inhibitory functions of BRCA1, BRCA1a, and BRCA1b proteins. *J Cell Physiol.* 2009 Jun;219(3):634-41.
- Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol.* 2001 Sep;18(9):1611-30. Review.
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science.* 1994;266(5182):66-71.
- Morris JR, Solomon E. BRCA1: BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Hum Mol Genet.* 2004;13(8):807-817.
- Muñoz MJ, de la Mata M, Kornblihtt AR. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem Sci.* 2010 Sep;35(9):497-504. Review.
- Orban TI, Olah E. Emerging roles of BRCA1 alternative splicing. *Mol Pathol.* 2003;56(4):191-7.
- Orban TI, Olah E. Expression profiles of BRCA1 splice variants in asynchronous and in G1/S synchronized tumor cell lines. *Biochem Biophys Res Commun.* 2001;280(1):32-8.
- Ouchi T. BRCA1 phosphorylation: biological consequences. *Cancer Biol Ther.* 2006;5(5):470-475.
- Paschalis A, Sharp A, Welti JC, Neeb A, Raj GV, Luo J, Plymate SR, de Bono JS. Alternative splicing in prostate cancer. *Nat Rev Clin Oncol.* 2018 Nov;15(11):663-675 Review.

- Paull TT, Cortez D, Bowers B, Elledge SJ, Gellert M. Direct DNA binding by Brca1. *Proc Natl Acad Sci USA*. 2001;98(11):6086-6091.
- Pickrell JK, Pai AA, Gilad Y, Pritchard JK. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet*. 2010 Dec 9;6(12):e1001236.
- Pohlreich P, Stribrna J, Zikan M, Novotny J. Splice variants of BRCA1 and BRCA2 genes identified in high-risk Czech breast and ovarian cancer families. 19th Meeting of European Association for Cancer Research Program/Proceedings, Budapest 2006;p260-p260.
- Pohlreich P, Zikan M, Stribrna J, Kleibl Z, Janatova M, Kotlas J, Zidovska J, Novotny J, Petruzelka L, Szabo C, Matous B. High proportion of recurrent germline mutations in the BRCA1 gene in breast and ovarian cancer patients from the Prague area. *Breast Cancer Res*. 2005;7(5):R728-36.
- Ponomarenko EA, Poverennaya EV, Ilgisonis EV, Pyatnitskiy MA, Kopylov AT, Zgoda VG, Lisitsa AV, Archakov AI. The Size of the Human Proteome: The Width and Depth. *Int J Anal Chem*. 2016;7436849. Review.
- Qin Y, Xu J, Aysola K, Oprea G, Reddy A, Matthews R, Okoli J, Cantor A, Grizzle WE, Partridge EE, Reddy ES, Landen C, Rao VN. BRCA1 proteins regulate growth of ovarian cancer cells by tethering Ubc9. *Am J Cancer Res*. 2012;2(5):540-8.
- Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. *Mol Genet Genomics*. 2017 Dec;292(6):1175-1195. Review.
- Rodríguez JA, Henderson BR. Identification of a functional nuclear export sequence in BRCA1. *J Biol Chem*. 2000;275(49):38589-96.
- Romero A, García-García F, López-Perolio I, Ruiz de Garibay G, García-Sáenz JA, Garre P, Ayllón P, Benito E, Dopazo J, Díaz-Rubio E, Caldés T, de la Hoya M. BRCA1 Alternative splicing landscape in breast tissue samples. *BMC Cancer*. 2015 Apr 3;15:219.
- Scherer WF, Syverton JT, Gey GO. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med*. 1953 May;97(5):695-710.
- Scholzová E, Malík R, Sevcík J, Kleibl Z. RNA regulation and cancer development. *Cancer Lett*. 2007 Feb 8;246(1-2):12-23. Review.

- Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, Livingston DM. Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*. 1997;88(2):265–275.
- Seo A, Steinberg-Shemer O, Unal S, Casadei S, Walsh T, Gumruk F, Shalev S, Shimamura A, Akarsu NA, Tamary H, King MC. Mechanism for survival of homozygous nonsense mutations in the tumor suppressor gene BRCA1. *Proc Natl Acad Sci USA*. 2018 May 15;115(20):5241-5246.
- Sevcik J, Falk M, Kleiblova P, Lhota F, Pohlreich P, Kleibl Z. In vitro analysis of population specific BRCA1 splicing variants (P- 507). *Eur J Cancer Suppl*. 2010;8(5):S130-S130
- Sevcik J, Falk M, Kleiblova P, Lhota F, Stefancikova L, Janatova M, Weiterova L, Lukasova E, Kozubek S, Pohlreich P, Kleibl Z. The BRCA1 alternative splicing variant $\Delta 14-15$ with an in-frame deletion of part of the regulatory serine-containing domain (SCD) impairs the DNA repair capacity in MCF-7 cells. *Cell Signal*. 2012;24(5):1023-30.
- Sevcik J, Falk M, Macurek L, Kleiblova P, Lhota F, Hojny J, Stefancikova L, Janatova M, Bartek J, Stribrna J, Hodny Z, Jezkova L, Pohlreich P, Kleibl Z. Expression of human BRCA1 $\Delta 17-19$ alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response. *Cell Signal*. 2013 May;25(5):1186-93.
- Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, King MC. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res*. 1996;6(11):1029-1049.
- Soule HD, Vazquez J, Long A, Albert S, Brennan M. A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst*. 1973 Nov;51(5):1409-16.
- Wang B, Hurov K, Hofmann K, Elledge SJ. NBA1, a new player in the BRCA1 A complex, is required for DNA damage resistance and checkpoint control. *Genes Dev*. 2009;23(6):729-739.
- Wang H, Shao N, Ding QM, Cui J, Reddy ES, Rao VN. BRCA1 proteins are transported to the nucleus in the absence of serum and splice variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases. *Oncogene*. 1997;15(2):143-57.
- Williams RS, Green R, Glover JN. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat Struct Biol*. 2001;8(10):838-842.

- Wilson CA, Payton MN, Elliott GS, Buaas FW, Cajulis EE, Grosshans D, Ramos L, Reese DM, Slamon DJ, Calzone FJ. Differential subcellular localization, expression and biological toxicity of BRCA1 and the splice variant BRCA1-delta11b. *Oncogene*. 1997 Jan 9;14(1):1-16.
- Wong AK, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. *J Biol Chem*. 1997;272(51):31941-31944.
- Wood MJ, Gait MJ, Yin H. RNA-targeted splice-correction therapy for neuromuscular disease. *Brain*. 2010;133(Pt 4):957-72.
- Xu J, Agyemang S, Qin Y, Aysola K, Giles M, Oprea G, O'Regan RM, Partridge EE, Harris-Hooker S, Rice VM, Reddy ES, Rao VN. A Novel Pathway that Links Caveolin-1 Down-Regulation to BRCA1 Dysfunction in Serous Epithelial Ovarian Cancer Cells. *Environ Health Perspect*. 2014;122(1): pii: 004.
- Xu J, Footman A, Qin Y, Aysola K, Black S, Reddy V, Singh K, Grizzle W, You S, Moellering D, Reddy ES, Fu Y, Rao VN. BRCA1 Mutation Leads to Deregulated Ubc9 Levels which Triggers Proliferation and Migration of Patient-Derived High Grade Serous Ovarian Cancer and Triple Negative Breast Cancer Cells. *Int J Chronic Dis Ther*. 2016;2(3):31-38.
- Xu J, Watkins T, Reddy A, Reddy ES, Rao VN. A novel mechanism whereby BRCA1/1a/1b fine tunes the dynamic complex interplay between SUMO-dependent/independent activities of Ubc9 on E2-induced ERalpha activation/repression and degradation in breast cancer cells. *Int J Oncol*. 2009 Apr;34(4):939-49.

PŘÍLOHY

Příloha 1: Seznam primerů navržených pro amplifikaci exon-exonových spojení v průběhu mPCR.

Název	Sekvence (5' → 3')	T _{an} (°C)	Délka (b)
B1 HTS e1F	CAGGAGGCCTTCACCCTC	55,7	18
B1 HTS e2F	CTATGCAGAAAATCTTAGAGTGTCCC	55,3	26
B1 HTS e3F	CTGTCTCCACAAAGTGTGACC	55,1	21
B1 HTS e5F	CACAGTGTCTTTTATGTAAGAATGATATAAC	54,6	31
B1 HTS e6F	ATTTGTGCTTTTCAGCTTGACAC	55,0	23
B1 HTS e7F	CTACAGAGTGAACCCGAAAATCC	55,2	23
B1 HTS e8F	CAACCTCAAAGACGTCTGTCTAC	55,6	24
B1 HTS e9F	CTGAAGATACCGTTAATAAGGCAAC	54,0	25
B1 HTS e10F	GGATGAAATCAGTTTGGATTCTGC	54,6	24
B1 HTS e11q F	CAGCTGAGAGGCATCCAG	53,9	18
B1 HTS e11F	CAAGAAGAGCAAAGCATGGATTG	54,7	23
B1 HTS e12F	TCCTCTCAGAGTGACATTTTAACC	54,1	24
B1 HTS e13F	CGAAATCCAGAACAAAGCACATC	55,0	23
B1 HTS e14F	TCTACCAGTAAAAATAAAGAACCAGGAG	54,7	28
B1 HTS e15F	GGAAACATCTTACTTGCCAAGGC	56,4	23
B1 HTS e16F	GTGGTGTCTGGCCTGAC	54,0	17
B1 HTS e17F	ACTGAAGAGACTACTCATGTTGTTATG	54,7	27
B1 HTS e18F	GGGAGGAAAATGGGTAGTTAGC	54,6	22
B1 HTS e19F	CCCAGTCTATTAAAGAAAGAAAAATGCTG	55,3	29
B1 HTS e20F	CAAAGCGAGCAAGAGAATCCC	55,9	21
B1 HTS e21F	CTATGGGCCCTTCACCAAC	54,3	19
B1 HTS e22F	GAAGGAGCTTTCATCATTACCC	55,7	23
B1 HTS e13alt F	CATGTTTTCTAAAAGGCCTAGAGAAC	54,3	26
B1 HTS e5q F	GAAGAAAGGGCCTTCACAGTG	55,4	21
B1 HTS e3R	GTGGAGACAGGTTCTTGATC	54,3	21
B1 HTS e5R	CCCTTTCTTCTGGTTGAGAAGTTTC	55,9	25
B1 HTS e6R	CAAGTTGACTAAATCTCGTACTTTCTTG	54,8	28
B1 HTS e7R	GTTATTTTCCTTTTTTGCAAATTATAGCTG	54,1	31
B1 HTS e8R	GTTAGAGAGTTGGACACTGAGAC	54,3	23
B1 HTS e9R	TGCCTTATTAACGGTATCTTCAGAAG	54,8	26
B1 HTS e10R	GAGGGGTGATTGTGAACAATTCTTG	54,5	25
B1 HTS e11R	GTATTTGTTACATCCGTCTCAGAAAATTC	55,2	29
B1 HTS e12R	CTGGGTGTGAGAGTGAAACAAG	55,3	22
B1 HTS e13R	GCTTTATCAGGTTATGTTGCATGG	54,7	24
B1 HTS e14R	AGGGTATTCACTACTTTTCTGTGAAG	54,6	26
B1 HTS e15R	CATCTAATGATGGGCATTTAGAAGGG	55,6	26
B1 HTS e16R	CTGATTCCAGATTCCAGGTAAGG	54,3	23
B1 HTS e17R	GGTGTTTTCTGGCAAACCTTGAC	55,0	23
B1 HTS e18R	TCAGTGTCCGTTACACAC	54,3	19
B1 HTS e19R	CAGCATTTTCTTTCTTTAATAGACTGGG	55,3	29
B1 HTS e20R	CATTGACCACATCTCCTCTGAC	54,7	22
B1 HTS e21R	GCCCATAGCAACAGATTCTAGC	55,8	23
B1 HTS e22R	CCACACAGCTGTACCATCC	54,4	19
B1 HTS e23R	CATCTGGCTGCACAACCAC	56,0	19
B1 HTS e24R	CACCACAGGTGCCTCAC	54,4	17
B1 HTS e13alt R	GGAGTTGTTCTTTTGCCATG	56,0	21

Příloha 2: Seznam primerů použitých pro RT-qPCR analýzu

Název amplikonu	Název primeru	Sekvence (5'→3')	T _{an} (° C)	Délka prim. (bp)	Délka ampl. (bp)
BRCA1 Δ3	B1 ex2 F	GAAAGAAATGGATTTATCTGCTCTTCGCG	58,9	29	109
	B1 del 3 R	GAGAAGTTTCAGCATGCAAAATCAGATG	57,6	28	
BRCA1 Δ3-5	B1 del3-5F	CTTAGAGTGTCCCATCTGGAGCC	58,1	23	104
	201R (e6)	CCAAACCTGTGTCAAGCTGAAAAGC	59,3	25	
BRCA1 Δ5	B1 del5F	GTGACCACATATTTTGCAAGAGCC	56,9	24	107
	B1SP 24R (e6)	CCAAACCTGTGTCAAGCTGAAAAG	57	24	
BRCA1 Δ9-10 (11q)	B1 del8-9	GTCTACATTGAATTGGCTGCTTGTG	57	25	148
	B1 ex 10a-11R	CAGATGCTGCTTCACCCTGATAC	57,4	23	
BRCA1 Δ9-10	B1 del8-9	GTCTACATTGAATTGGCTGCTTGTG	57	25	204
	B1SP 11R	CTCATGCTGTAATGAGCTGGC	55,7	21	
BRCA1 Δ9-11	B1 del9-11 F	GTCTGTCTACATTGAATTGGGTGAAGC	58,1	27	109
	B1 ex12 R	CTGAGTGGTTAAAATGTCACTCTGAGAGG	58,8	29	
BRCA1 Δ11	B1 del11F	GTTTGGATTCTGCAAAAAGGGTGAAG	57,8	27	110
	B1 ex12 R	CTGAGTGGTTAAAATGTCACTCTGAGAGG	58,8	29	
BRCA1 11q	B1 ex11kr-12 F	ATCCAGAAAAGTATCAGG GTGAAGC	56,7	25	107
	B1 ex12 R	CTGAGTGGTTAAAATGTCACTCTGAGAGG	58,8	29	
BRCA1 e13A	B1 ex 13 F	GGGAGCCAGCCTTCTAACAGC	59,3	21	123
	B1 alt 13 R1	GTTGTTTCCTTTGGCCATGTATATGCG	58,9	26	
BRCA1 IRIS	B1 Iris 01f	GATTCAAACCTTAGGTATTGGAACCAGG	55,8	27	105
	B1 Iris 02r	GGAAATTTGTAAAATGTGCTCCCCA	56,3	25	
BRCA1 e7	B1 ex7 F	ATGCAAACAGCTATAATTTTGCAAAAAGG	56,9	30	137
	B1 ex7 R	GGAAGGATTTTCGGGTTCACCTCTG	57,6	24	
BRCA1 e23-24	B1 ex23 F	TCCACCCAATTGTGGTTGTGC	57,9	21	130
	B1 ex 24R (cds)	CACTGGTAGAGTGCTACACTGTCC	58,4	24	
CHEK2 e7	CHK11f	CTTGTGGTTTTCTCTTGGA	54,3	21	213
	CHK12r	GATGAGAAAGGCAAGCCTACA	54,3	21	
UBC	UBC-01f	GATCGCTGTGATCGTCACTTG	55,8	21	169
	UBC-02r	GTTTTCCAGCAAAGATCAGCCT	55,6	22	
ATP5B	ATP5B-01f	TGAAGAAGCTGTGGCAAAAGC	56,2	21	106
	ATP5B-02r	GAAGCTTTTGGGTAGGGGC	56,0	21	
GAPDH	hGAPDH-01f	GGTGAAGGTCGGAGTCAACGG	59,3	21	231
	hGAPDH-02r	CGCTCCTGGAAGATGGTGATGG	59,2	22	
BRCA1 e2-e6	B1 ex2 F	GAAAGAAATGGATTTATCTGCTCTTCGCG	58,9	29	305
	201R (e6)	CCAAACCTGTGTCAAGCTGAAAAGC	59,3	25	

Příloha 3: Procentuální zastoupení variant využívající volná sestřihová místa ve vyšetřovaných skupinách

Název varianty	HGVS nomenklatura	Biotyp	MCF7	MIX cDNA	Kontroly Leukocyty	BC Leukocyty	BRCA1 mut. Leukocyty	Kontroly Mama	BC Mama	BRCA1 mut. Mama	Kontroly tuková tk.	BC tuková tk.	BRCA1 mut. tuková tk.
e1A (wt)	exon 1	wt	51,4 %	44,2 %	62,8 %	63,4 %	61,9 %	54,6 %	55,8 %	57,7 %	56,1 %	54,4 %	55,0 %
1Aq	c.-25_-20del6	SDSΔ	48,6 %	55,8 %	37,2 %	36,6 %	38,1 %	45,4 %	44,2 %	42,3 %	43,9 %	45,6 %	45,0 %
e2 (wt)	exon 2	wt	96,4 %	93,2 %	98,6 %	99,1 %	97,9 %	98,8 %	99,2 %	99,7 %	98,7 %	98,2 %	98,5 %
2p	c.-19_-7del13	SASΔ	3,6 %	6,8 %	1,4 %	0,9 %	2,1 %	1,2 %	0,8 %	0,3 %	1,3 %	1,8 %	1,5 %
e6 (wt)	exon 6	wt	100 %	99,7 %	99,5 %	99,7 %	99,1 %	99,7 %	99,8 %	100 %	99,8 %	100 %	99,9 %
6q	c.293_301del9	SASΔ	0,0 %	0,3 %	0,5 %	0,3 %	0,9 %	0,3 %	0,2 %	0,0 %	0,2 %	0,0 %	0,1 %
e8 (wt)	exon 8	wt	85,7 %	73,6 %	71,9 %	76,7 %	70,4 %	72,6 %	79,9 %	72,7 %	74,0 %	73,7 %	74,8 %
8p	c.442_444del3	SASΔ	14,3 %	26,4 %	28,1 %	23,3 %	29,6 %	27,4 %	20,1 %	27,3 %	26,0 %	26,3 %	25,2 %
e10 (wt)	exon 10	wt	32,7 %	86,5 %	80,5 %	69,5 %	95,0 %	93,3 %	70,8 %	91,5 %	92,5 %	96,3 %	93,8 %
10a	c.594-21_594-1 ins21	SAS▼	67,3 %	13,5 %	19,5 %	30,5 %	5,0 %	6,7 %	29,2 %	8,5 %	7,5 %	3,7 %	6,2 %
e13 (wt)	exon 13	wt	94,4 %	89,5 %	89,0 %	88,8 %	87,6 %	89,2 %	92,3 %	91,8 %	90,3 %	94,6 %	90,4 %
13p	c.4186_4188del3	SASΔ	5,6 %	10,5 %	11,0 %	11,2 %	12,4 %	10,8 %	7,7 %	8,2 %	9,7 %	5,4 %	9,6 %
e14 (wt)	exon 14	wt	63,8 %	70,6 %	73,4 %	74,2 %	84,0 %	81,2 %	88,1 %	88,8 %	82,9 %	78,6 %	88,9 %
14p	c.4358_4360del3	SASΔ	36,2 %	29,4 %	26,6 %	25,8 %	16,0 %	18,8 %	11,9 %	11,2 %	17,1 %	21,4 %	11,1 %
e23 (wt)	exon 23	wt	99,1 %	99,2 %	99,5 %	100 %	99,4 %	98,8 %	97,8 %	98,8 %	98,2 %	98,7 %	100 %
23a	c.5407-9_5407-1 ins9	SAS▼	0,9 %	0,8 %	0,5 %	0,0 %	0,6 %	1,2 %	2,2 %	1,2 %	1,8 %	1,3 %	0,0 %
Δ2	c.-19_80del199	CΔ	50,0 %	44,2 %	92,3 %	91,3 %	93,0 %	87,6 %	94,5 %	98,0 %	89,0 %	94,1 %	69,2 %
1Aq, Δ2	c.-25_80del105	SDSΔ + CΔ	50,0 %	55,8 %	7,7 %	8,7 %	7,0 %	12,4 %	5,5 %	2,0 %	11,0 %	5,9 %	30,8 %
Δ2_3	c.-19_134del153	mCΔ	-	50,0 %	-	0,0 %	0,0 %	72,0 %	88,7 %	-	100 %	-	100 %
1Aq, Δ2_3	c.-25_134del159	SDSΔ + mCΔ	-	50,0 %	-	100 %	100 %	28,0 %	11,3 %	-	0,0 %	-	0,0 %
Δ2_5	c.-19_212del231	mCΔ	100 %	60,4 %	100 %	-	-	97,4 %	100 %	0,0 %	-	-	-
1Aq, Δ2_5	c.-25_212del237	SDSΔ + mCΔ	0,0 %	22,0 %	0,0 %	-	-	2,6 %	0,0 %	100 %	-	-	-
1Aq, Δ2_5, 6p	c.-25_282del307	SDSΔ + mCΔ + SASΔ	0,0 %	17,6 %	0,0 %	-	-	0,0 %	0,0 %	0,0 %	-	-	-
Δ2_7, 8p	c.-19_444del463	mCΔ + SASΔ	-	100 %	100 %	83,3 %	-	0,0 %	-	-	-	-	-
1Aq, Δ2_7, 8p	c.-25_444del469	SDSΔ + mCΔ + SASΔ	-	0,0 %	0,0 %	16,7 %	-	0,0 %	-	-	-	-	-
1Aq, Δ2_7	c.-25_441del466	SDSΔ + mCΔ	-	0,0 %	0,0 %	0,0 %	-	100 %	-	-	-	-	-

Název varianty	HGVS nomenklatura	Biotyp	MCF7	MIX cDNA	Kontroly Leukocyty	C50 Leukocyty	BRCA1 mut. Leukocyty	Kontroly Mama	C50 Mama	BRCA1 mut. Mama	Kontroly tuková tk.	C50 tuková tk.	BRCA1 mut. tuková tk.
Δ2_10	c.-19_670del689	mCΔ	-	16,7 %	-	-	0,0 %	-	-	-	-	-	-
1Aq, Δ2_10	c.-25_670del695	SDSΔ + mCΔ	-	83,3 %	-	-	100 %	-	-	-	-	-	-
Δ2_17	c.-19_5074del5093	mCΔ	0,0 %	68,2 %	-	-	-	-	-	-	100 %	-	-
1Aq, Δ2_17	c.-25_5074del5099	SDSΔ + mCΔ	100 %	31,8 %	-	-	-	-	-	-	0,0 %	-	-
Δ2_19	c.-19_5193del5212	mCΔ	-	-	-	100 %	-	-	77,2 %	-	-	100 %	-
1Aq, Δ2_19	c.-25_5193del5218	SDSΔ + mCΔ	-	-	-	0,0 %	-	-	22,8 %	-	-	0,0 %	-
Δ5_6	c.135_301del167	mCΔ	-	-	-	-	0,0 %	-	0,0 %	100 %	-	-	-
Δ5_6, 7p	c.135_307del173	mCΔ + SASΔ	-	-	-	-	100 %	-	100 %	0,0 %	-	-	-
Δ5_7	c.135_441del307	mCΔ	-	100 %	0,0 %	-	-	-	-	-	-	-	-
Δ5_7, 8p	c.135_444del310	mCΔ + SASΔ	-	0,0 %	100 %	-	-	-	-	-	-	-	-
Δ11_12	c.671_4185del3515	mCΔ	-	0,0 %	-	-	-	100 %	100 %	-	-	-	-
Δ11_12, 13p	c.671_4188del3518	mCΔ + SASΔ	-	100 %	-	-	-	0,0 %	0,0 %	-	-	-	-
Δ13	c.4186_4357del172	CΔ	50,0 %	82,1 %	77,8 %	100 %	100 %	85,7 %	85,7 %	100 %	100 %	100 %	100 %
Δ13, 14p	c.4186_4360del175	CΔ + SASΔ	50,0 %	17,9 %	22,2 %	0,0 %	0,0 %	14,3 %	14,3 %	0,0 %	0,0 %	0,0 %	0,0 %
▼13A,14	c.4358-2785_4358-2720ins66	C▼	49,6 %	37,9 %	96,8 %	92,1 %	97,8 %	92,1 %	99,0 %	98,1 %	96,3 %	98,0 %	99,5 %
▼13A,14 p	c.4358-2785_4358-2720ins66 + c.4358_4360del3	C▼ + SASΔ	50,4 %	62,1 %	3,2 %	7,9 %	2,2 %	7,9 %	1,0 %	1,9 %	3,7 %	2,0 %	0,5 %

Příloha 4 – Plný text manuskriptu: Hojny J, Zemankova P, Lhota F, Sevcik J, Stranecky V, Hartmannova H, Hodanova K, Mestak O, Pavlista D, Janatova M, Soukupova J, Vocka M, Kleibl Z, Kleiblova P. **Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model.** *Gene*. 2017 Dec 30;637:41-49. doi:10.1016/j.gene.2017.09.025.

Příloha 5 – Plný text manuskriptu: Sevcik J, Falk M, Macurek L, Kleiblova P, Lhota F, Hojny J, Stefancikova L, Janatova M, Bartek J, Stribrna J, Hodny Z, Jezkova L, Pohlreich P, Kleibl Z. **Expression of human BRCA1 Δ 17-19 alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response.** Cell Signal. 2013 May;25(5):1186-93. doi:10.1016/j.cellsig.2013.02.008